Karlsruher Institut für Technologie

ETP-KA/2025-6

Classical and Quantum Machine Learning for Anomaly Detection at the CMS Experiment and Beyond

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

angenommene

Dissertation

von

M.Sc. Aritra Bal

aus Kolkata (Indien)

Tag der mündlichen Prüfung: 11 Juli 2025
1. Referent: Prof. Dr. Markus Klute
2. Korreferent: Priv. Doz. Dr. Roger Wolf
3. Betreuer: Dr. Benedikt Maier

Declaration:

I hereby declare that I have completed this work independently and have not used any sources or aids other than those cited in the text. All content taken from other works, whether quoted verbatim or used with modifications, has been appropriately cited. If writing aids such as Large Language Models (LLMs) have been used in this thesis, I confirm that their output was derived from my own original work, and the sole function of the tool was to correct grammar and language wherever applicable. I take full responsibility for all erroneous content, incorrect references and plagiarism arising from the usage of such tools.

Karlsruhe, July 30, 2025

. . . . . .
Aritra Bal

Abstract

This thesis presents two novel methodologies for anomaly detection in particle collision events, utilising both classical and quantum machine learning techniques. The first approach employs an unsupervised, data-driven variational autoencoder (VAE), trained on collision data recorded by the CMS detector at the Large Hadron Collider during Run 2, with a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. Events containing at least two large-radius jets are analysed, and the VAE-derived anomaly score is subsequently decorrelated from the dijet invariant mass (m_{jj}) through a novel machine learning-based approach. A systematic search for potential anomalies that deviate from the expected smoothly falling background distribution is performed in the m_{jj} range from 1.8 to 6 TeV. No significant excess above the Standard Model (SM) expectation is observed, with the highest observed local significance being 2.3σ at 4.9 TeV. This method is subsequently used to derive exclusion limits on a large number of exotic signal models derived from Beyond the Standard Model (BSM) physics.

The second approach introduces a novel quantum machine learning framework, named *1P1Q*, for encoding jet kinematic features onto two-level quantum bits (*qubits*). Using this encoding strategy, a quantum autoencoder is trained to capture the underlying distribution of jets initiated by a light quark or gluon, and subsequently generate anomaly scores for exotic jets with a more complex jet substructure. The developed quantum data encoding is also validated in a jet classification task, for distinguishing jets arising from the decay of a top quark, from those initiated by a light quark or gluon. For both the anomaly detection and supervised classification tasks, the *1P1Q* method is demonstrated to achieve signal identification and separation performance comparable to or surpassing current state-of-the-art classical machine learning algorithms.

Contents

List of Abbreviations

AE Autoencoder.

ALICE A Large Ion Collider.

AOJ Aspen Open Jets.

APD Avalanche Photo Diode.

ASIC Application Specific Integrated Circuit.

ATLAS A Toroidal LHC ApparatuS.

AUC Area Under the Curve.

BCE Binary Cross Entropy.

BSM Beyond the Standard Model.

CATHODE Classifying Anomalies THrough Outer Density Estimation.

CERN European Organisation for Nuclear Research.

CMS Compact Muon Solenoid.

CNN Convolutional Neural Network.

CP Charge-Parity.

CR Control Region.

CWoLa Classification Wit**h**o**u**t L**a**bels.

DCB Double Crystal Ball.

DNN Deep Neural Network.

ECAL Electromagnetic Calorimeter.

ELBO Evidence Lower BOund.

EPR Einstein-Podolsky-Rosen.

ETP Institute of Experimental Particle Physics.

FPGA Field Programmable Gate Array.

FPR False Positive Rate.

GAN Generative Adversarial Network.

HEP High Energy Physics.

HL-LHC High Luminosity LHC.

JER Jet Energy Resolution.

JES Jet Energy Scale.

KIT Karlsruhe Institute of Technology.

LHC Large Hadron Collider.

LHCb Large Hadron Collider beauty.

MC Monte-Carlo.

MET Missing Transverse Energy.

ML Machine Learning.

MLP Multi-Layer Perceptron.

NISQ Noisy Intermediate-Scale Quantum.

NN Neural Network.

PF Particle Flow.

PFCand Particle Flow Candidate.

PQC Parameterised Quantum Circuit.

QAE Quantum Autoencoder.

QCD Quantum Chromodynamics.

QML Quantum Machine Learning.

QNN Quantum Neural Network.

QR Quantile Regression.

QUAK QUasi Anomalous Knowledge.

ROC Receiver Operating Characteristics.

SGD Stochastic Gradient Descent.

SM Standard Model.

SR Signal Region.

TNT Tag 'N Train.

TPR True Positive Rate.

VAE Variational Autoencoder.

VAE-QR Variational Autoencoder-Quantile Regression.

VEV Vacuum Expectation Value.

VQC Variational Quantum Classifier.

XOR Exclusive-OR.

List of Figures

List of Tables

Part I.

Introduction

1. Introduction

1.1. Motivation

The Standard Model (SM) of particle physics [1, 2, 3] stands as one of the most successful scientific theories developed to date. It provides a comprehensive framework for understanding all known fundamental particles and their interactions, and its predictions have been confirmed by several experiments with remarkable precision. The discovery of the Higgs boson at the LHC in 2012 [4, 5, 6] completed the particle content of the SM and further validated its accuracy. However, despite its successes, the SM is widely regarded as incomplete. It fails to explain several critical phenomena observed in the universe, such as the nature of dark matter, the origin of the matter-antimatter asymmetry, and the hierarchy problem, among others. Moreover, the SM does not incorporate gravity and remains silent on the quantisation of spacetime.

The ATLAS and CMS experiments at the LHC have conducted an extensive program of measurements and searches, over the past decade. During Run 2 of the LHC, which delivered approximately 138 $\text{fb}^{-1}$of proton-proton collision data at $\sqrt{s} = 13$ TeV, the SM was tested with unprecedented precision. Nonetheless, no clear evidence for Beyond the Standard Model (BSM) physics has emerged from these efforts. Over the past decade, both collaborations have undertaken hundreds of dedicated searches targeting a wide range of scenarios not explainable by the SM, including but not limited to supersymmetry, new heavy resonances and dark matter candidates. Until now, all observations have turned out to be consistent with SM expectations.

The absence of discovery, however, does not imply the absence of new physics. Rather, it calls for a paradigm shift in the way one searches for it. Traditionally, BSM searches have relied on specific theoretical models, with the expected final states being already known. Such a strategy, though powerful, is inherently limited. One runs the risk of missing signals that do not conform to predefined assumptions. In this context, model-agnostic approaches that make the fewest possible assumptions about the nature of new physics become increasingly important. These approaches aim to identify deviations from the SM in a data-driven manner, offering a complementary path towards discovery.

This thesis presents the first fully data-driven and model-agnostic search for BSM physics in dijet final states on data collected by the CMS detector during Run 2, using advanced ML algorithms for anomaly detection. The search strategy makes minimal assumptions about the signals that it searches for, requiring only generic resonances decaying to two jets in the final state. This ML-based search is implemented in two stages. First, a Variational Autoencoder (VAE) is trained on a signal-depleted control region dominated by SM jets to learn the underlying probability distribution. It is then applied to events in the signal region to calculate an anomaly score for each event (containing two jets) based on a reconstruction loss. Second, a DNN is trained to decorrelate the anomaly score from the invariant mass of the

dijet system using a Quantile Regression (QR), thereby mitigating artificial sculpting of the background distribution. A search for resonant signals that deviate from the SM expectation of a smoothly falling background is then performed over the m_{jj} spectrum, to look for localised excesses indicative of a potential signal. No significant deviation from the SM was observed in this analysis. Thereafter, upper limits were set on several benchmark signal models, many of these for the first time ever at the CMS Experiment. Importantly, the method remains agnostic to the specific properties of the signal, making it a flexible tool for future investigations using arbitrary signal models.

To complement the classical approach, this thesis also explores the use of Quantum Machine Learning (QML) in the search for new physics. Quantum computing, particularly in the Noisy Intermediate-Scale Quantum (NISQ) era, offers novel computational paradigms that may provide advantages in certain learning tasks. Recent studies [7, 8] suggest that quantum algorithms can in general capture complex structures in high-dimensional data more efficiently than their classical counterparts, and exploit in specific the inherently geometric nature of particle physics data. In this thesis a new quantum data encoding scheme, called *1P1Q* that maps the kinematic features of jet constituents directly onto qubits is introduced. Each jet is thus represented as a tensor product of Hilbert spaces, easily visualisable as points on the Bloch sphere, allowing for a compact and lossless representation without the need for classical preprocessing or dimensionality reduction.

Using this encoding, a quantum autoencoder is trained on simulated QCD jets and then applied to jets from resonant decays. The resulting anomaly scores demonstrate strong discriminatory power, with the quantum model achieving performance comparable to, or exceeding, state-of-the-art classical benchmarks while using significantly fewer trainable parameters. Additionally, this encoding is shown to be effective in a supervised classification task distinguishing top quark jets from jets initiated by a light quark or gluon, further highlighting its versatility. To demonstrate its robustness in realistic conditions, the quantum encoding and algorithms are applied to collision data recorded by the CMS detector in 2016, showing that it remains stable and effective even when deployed on experimental data with all associated detector effects and uncertainties.

In summary, this thesis contributes two novel methodologies for BSM searches: a fully data-driven, unsupervised classical machine learning pipeline, and a complementary quantum approach leveraging the power of variational quantum circuits. Both methods aim to broaden the discovery potential of the LHC by enabling flexible, model-agnostic analyses that are sensitive to a wide range of possible new physics signatures.

Part II.

Introduction to Theoretical and Experimental High Energy Physics

2. Foundations of High Energy Physics

The study of elementary particles and their interactions represents one of the most profound pursuits in modern physics. This chapter establishes the theoretical and experimental foundations necessary for understanding the research presented in this thesis. First, the mathematical framework of the SM of particle physics is explored, including its constituent particles, fundamental interactions, and the symmetry principles that govern their behaviour. The Higgs mechanism, which explains how gauge bosons and fermions acquire mass through spontaneous symmetry breaking and the existence of the Higgs field [9, 10, 11], is shortly described. Thereafter, the discussion transitions to experimental HEP, which is the primary concern of this thesis.

2.1. Theoretical Introduction to the Standard Model (SM)

The SM is a relativistic quantum field theory that describes the fundamental constituents of matter and the forces governing their interactions, excluding gravity [1, 2, 3]. Developed progressively throughout the latter half of the 20th century, this theoretical framework represents the culmination of decades of collaborative scientific effort to unify the electromagnetic and weak interactions [12] and to categorise the diverse array of subatomic particles discovered in accelerator experiments [13, 14]. The SM unifies three of the four known fundamental interactions, these being the electromagnetic, weak, and strong interactions, within a consistent mathematical structure based on gauge symmetry principles [15]. Despite its incompleteness, most notably in its inability to incorporate gravitational interactions and explain phenomena such as dark matter, the SM has demonstrated remarkable success, with its theoretical predictions being verified through numerous high-precision experimental measurements, ultimately culminating in the discovery of the Higgs boson in 2012 [4, 5, 6] at the Large Hadron Collider (LHC).

2.1.1. Overview of the Elementary Particles

The SM organises elementary particles into two main categories: spin-$\frac{1}{2}$ fermions, which constitute matter, and integer-spin bosons, which mediate forces. Fermions are further subdivided into quarks, which participate in strong interactions, and leptons, which do not. Both quarks and leptons are arranged in three generations with identical quantum numbers but increasing masses. Each generation of quarks contains an up-type quark (up, charm, top) with charge $+\frac{2}{3}e$ and a down-type quark (down, strange, bottom) with charge $-\frac{1}{3}e$ [16, 13, 14]. Similarly, each lepton generation consists of a charged lepton (electron, muon, tau) with charge $-e$ and a corresponding neutral neutrino (electron neutrino, muon neutrino, tau neutrino) [17, 18]. The force-mediators include the massless photon mediating electromagnetism, the massive

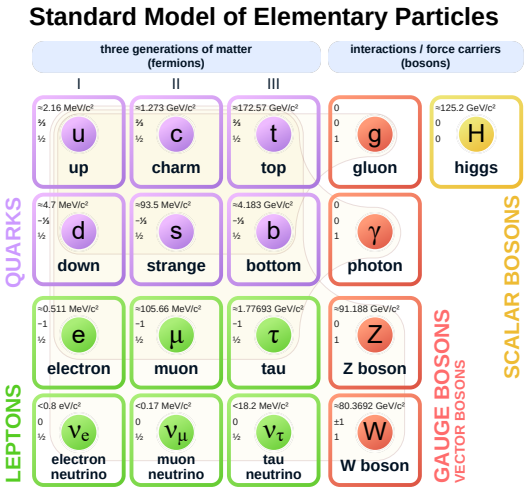


Figure 2.1.: The Standard Model of Particle Physics [22].

$W^{\pm}$ and Z^0 bosons responsible for weak interactions, and eight massless gluons that carry the strong force [19, 20, 21]. The framework of the SM is completed by the scalar Higgs boson, which imparts mass to elementary particles [4, 5, 9, 10, 11]. As of today, the SM successfully describes most observed particle physics phenomena, though important questions remain unresolved.

The particles described by the SM are characterised by a set of fundamental attributes: mass, electric charge, colour (for quarks), and intrinsic spin. Additionally, in the electroweak sector, fermions possess the fundamental attributes of weak isospin (T_3) and weak hypercharge (Y), which determine how they transform under the $SU(2)_L \times U(1)_Y$ gauge symmetry. Left-handed fermions transform as doublets under $SU(2)_L$ with $T_3 = \pm 1/2$, while right-handed fermions are singlets with $T_3 = 0$. These electroweak attributes are related to the electric charge Q by the formula $Q = T_3 + Y/2$, reflecting the unification of electromagnetic and weak interactions [1, 2, 3]. In the strong interaction sector, quarks and gluons carry colour charge and transform under the $SU(3)_C$ symmetry, with quarks in the fundamental representation and gluons in the adjoint representation. This sector exhibits two remarkable properties: asymptotic freedom at high energies, where the strong coupling becomes weak, and colour confinement at low energies, which prevents the observation of isolated coloured particles [20, 21]. The Higgs sector introduces the final piece of the SM through a complex scalar doublet that acquires a non-zero Vacuum Expectation Value (VEV), spontaneously breaking the electroweak symmetry and generating masses for the $W^{\pm}$ and Z bosons while leaving the photon massless [9, 10, 11]. The mechanism also provides mass to fermions through Yukawa couplings to the Higgs field. In the SM, quarks exhibit a wide range of masses depending on their generation, from a few MeV for the up and down quarks up to about 173 GeV for the top quark, whereas charged leptons (electron, muon, tau) have masses spanning from 0.511 MeV for the electron to 1.777 GeV for the tau lepton. The neutrinos are electrically neutral leptons whose masses are exceedingly small and as the SM predicts, were long thought to be massless. This was until a series of landmark experiments demonstrated the phenomenon of neutrino oscillation [17, 18] and proved that they must possess a small non-zero mass. Despite this discovery, the absolute neutrino mass scale remains unknown, and experiments cannot yet distinguish whether the mass eigenstate predominantly associated with the electron neutrino is the lightest (normal hierarchy) or the heaviest (inverted hierarchy), both being permitted by theory [23] with the

former being slightly preferred [24]. Resolving this ordering is crucial for understanding how neutrino masses arise, and as a result, this stands as one of the major challenges for the SM, indicating a need for theories that look beyond.

2.1.2. Mathematical Structure of the Standard Model

The SM is a relativistic quantum field theory based on the gauge group

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

which unifies our current understanding of the strong, weak, and electromagnetic interactions into a single theoretical framework. Its Lagrangian density can be divided into four main parts:

$$\mathcal{L}_{\mathrm{SM}} = \mathcal{L}_{\mathrm{gauge}} + \mathcal{L}_{\mathrm{fermion}} + \mathcal{L}_{\mathrm{Higgs}} + \mathcal{L}_{\mathrm{Yukawa}}.$$

Each term is shortly described below:

- **Gauge Fields ($\mathcal{L}_{\mathrm{gauge}}$):**

$$\mathcal{L}_{\mathrm{gauge}} = -\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} \; - \; \frac{1}{4} W^i_{\mu\nu} W^{i\mu\nu} \; - \; \frac{1}{4} B_{\mu\nu} B^{\mu\nu}.$$

  This term describes the gauge bosons of the SM. Specifically:

  - **Gluons**: These arise from the $\mathrm{SU(3)_C}$ gauge group and come in eight varieties, corresponding to the adjoint representation indexed by $a = 1, \ldots, 8$. Gluons are massless and mediate the strong interaction among quarks.
  - **Weak bosons**: These originate from the $\mathrm{SU(2)_L}$ gauge group, with fields indexed by $i = 1, 2, 3$. In the absence of the Higgs mechanism, they would remain massless like the gluons. However, a direct mass term for gauge bosons is forbidden by local gauge symmetry. Instead, once the Higgs field acquires a nonzero VEV (see Section 2.1.3), the electroweak symmetry $\mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$ is spontaneously broken, in what is known as the Higgs Mechanism. As a result, these initially massless bosons acquire their mass and are known as the massive W^+, W^-, and Z^0 bosons, which mediate the weak interaction.
  - **Hypercharge field and the photon**: The remaining gauge symmetry, identified with $\mathrm{U(1)_Y}$, mixes with the $\mathrm{SU(2)_L}$ bosons to produce one additional massless particle, the photon, which mediates the electromagnetic force.

- **Fermionic Term ($\mathcal{L}_{\mathrm{fermion}}$):**

$$\mathcal{L}_{\mathrm{fermion}} = \sum_{\psi} i\, \bar{\psi}\, \gamma^\mu\, D_\mu\, \psi,$$

  where the sum runs over all quark and lepton fields, denoted collectively by ψ. The operator $D_\mu = \partial_\mu - i g_s T^a G^a_\mu - i g \frac{\sigma^i}{2} W^i_\mu - i g' Y B_\mu$ is the covariant derivative, which ensures that fermions couple appropriately to each gauge field. Here, g_s, g, and g' are the coupling constants for the $SU(3)_\mathrm{C}$, $SU(2)_\mathrm{L}$ and $U(1)_\mathrm{Y}$ gauge groups. The generators T^a (which can be represented using the Gell-Mann matrices) and $\sigma^i/2$ (represented by Pauli matrices) show how quarks and leptons transform under the $\mathrm{SU(3)_C}$ and $\mathrm{SU(2)_L}$ groups,

respectively, with Y denoting the weak hypercharge of the field. As discussed in the gauge field sector, before electroweak symmetry breaking, the $\mathrm{SU(2)_L}$ and $\mathrm{U(1)_Y}$ gauge fields are initially massless. As mentioned above, once the Higgs field obtains its VEV, the $\mathrm{SU(2)_L \times U(1)_Y}$ symmetry is spontaneously broken, and the fields mix to form the physical $W^{\pm}$ and Z^0 bosons, as well as the photon. As can be seen, the $W^{\pm}$ bosons carry electric charge. On the other hand, the Z^0 boson and the photon which emerge through a mixing governed by the weak mixing angle, θ_W are neutral. The angle θ_W quantifies the mixing between the neutral weak field (W^3_μ) and the weak hypercharge field (B_μ), which combine to form the Z^0 boson and the photon. Fermions interact with each of these physical gauge bosons consistently, according to their charge and weak isospin.

A key feature of the electroweak sector is that it violates parity (P) symmetry (replacing $\vec{x} \to -\vec{x}$). This is because left-handed fermions (and right-handed antifermions) transform as doublets under the $\mathrm{SU(2)_L}$ symmetry, while right-handed fermions (and left-handed antifermions) are singlets. Consequently, the charged weak bosons $W^{\pm}$ couple only to left-handed particles (and right-handed antiparticles) respectively, which implies that the weak interaction **distinguishes between** the two chiralities. These implications are significant, in the sense that mirror images of certain processes may occur with very different probabilities or not at all, leading to observable phenomena such as the left-handed nature of neutrinos.

- **Higgs Sector ($\mathcal{L}_{\text{Higgs}}$):**

$$\mathcal{L}_{\text{Higgs}} = |D_\mu \phi|^2 \; - \; V(\phi), \quad \text{where } \; V(\phi) = \mu^2 \, \phi^\dagger \phi \; + \; \lambda \, (\phi^\dagger \phi)^2.$$

The Higgs field ϕ is an $\mathrm{SU(2)_L}$ doublet, through which electroweak symmetry breaking occurs when $\mu^2 < 0$. The phenomenon of spontaneous symmetry breaking gives masses to the $W^{\pm}$ and Z^0 bosons (and, indirectly, to fermions through Yukawa interactions). The self-interaction governed by the potential $V(\phi)$ leads to a non-zero VEV for ϕ. The Higgs boson (the so-called quantum of the Higgs field) is therefore physical, and is produced and can be detected at collider experiments. The parameter λ is directly related to the physical Higgs boson mass once symmetry breaking is accounted for.

- **Yukawa Interactions ($\mathcal{L}_{\text{Yukawa}}$):**

$$\mathcal{L}_{\text{Yukawa}} = - \, y^d_{ij} \, \bar{Q}^i_L \, \phi \, d^j_R \; - \; y^u_{ij} \, \bar{Q}^i_L \, \tilde{\phi} \, u^j_R \; - \; y^e_{ij} \, \bar{L}^i_L \, \phi \, e^j_R \; + \; \text{h.c.},$$

with $\tilde{\phi} = i \, \sigma_2 \, \phi^*$. These terms describe how quarks and charged leptons acquire mass through interactions with the Higgs field. When the Higgs field obtains a VEV, these couplings become mass terms proportional to y_{ij}. The matrices y_{ij} determine the mass hierarchies among the different generations of fermions, and critically, cannot be predicted by theory alone. The large differences in fermion masses stem from similarly large differences in these Yukawa coupling strengths, as observed experimentally.

In the SM, local gauge invariance is preserved once all the pieces are included, ensuring the cancellation of anomalies that could otherwise compromise the mathematical consistency. Moreover, the theory is renormalisable, meaning that it remains predictive at higher energies. However, a key limitation is that the SM is unable to account for the phenomenon of gravity in its mathematical formulation. The three fundamental classes of couplings present in the SM are:

- *Gauge couplings* (g_s, g, g'): Set the strength of QCD and electroweak interactions
- *Yukawa couplings* (y_{ij}): Determine the masses and mixing patterns of the fermions
- *Higgs self-coupling* (λ): Fixes the mass and dynamics of the Higgs boson

Despite describing a remarkable array of phenomena up to energies probed by current experiments, the SM does not address all fundamental questions. In particular, it offers no explanation for the large spectrum of Yukawa couplings, it leaves neutrino mass generation beyond the Dirac framework unresolved, and it excludes a quantum theory of gravity. These open issues motivate many experimental searches and theoretical extensions, ranging from direct searches for additional particles at colliders to precision measurements aiming to probe new physics through subtle deviations from SM predictions.

2.1.3. Spontaneous Symmetry Breaking and the Higgs Mechanism

In the SM, bosons are the mediators of the fundamental interactions, each arising from local gauge symmetries in the theory. The massless *photon* (γ) carries the electromagnetic force, coupling to electric charge. The eight massless *gluons* (g) mediate the strong interaction, themselves carrying colour charge and hence self-interacting, and the massive $W^{\pm}$ and Z^0 bosons mediate the weak force. The $W^{\pm}$ and Z^0 bosons were discovered at the European Organisation for Nuclear Research (CERN) in 1983, confirming the electroweak theory [25, 26, 19]. The last major missing piece in the SM was the *Higgs boson*, a spin-0 scalar field responsible for spontaneous electroweak symmetry breaking. Discovered simultaneously by the A Toroidal LHC ApparatuS (ATLAS) and CMS Collaborations [4, 5] at the LHC in 2012, its existence verified the last major theoretical component of the SM, providing the mechanism, as described below, by which elementary particles acquire mass without breaking gauge invariance.

The Higgs mechanism predicts that the electroweak gauge group $\mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}$ is spontaneously broken down to the electromagnetic subgroup $\mathrm{U}(1)_\mathrm{EM}$ by a scalar field ϕ that acquires a nonzero VEV [1, 2, 3, 9, 10, 11]. Mathematically, one may express the Higgs field as a complex doublet under $\mathrm{SU}(2)_\mathrm{L}$, and therefore choose its potential such that the field settles into a minimum away from the origin [15]. This results in the gauge bosons $W^{\pm}$ and Z^0 acquiring mass, as the would-be Goldstone bosons become their longitudinal components, while the photon γ remains massless [2, 3]. For fermions, mass arises through *Yukawa couplings*, with each fermion coupling to the Higgs field through an interaction of the form $\bar{\psi}_L \, \phi \, \psi_R$. Within the framework of the SM, neutrinos remain massless due to the absence of a right-handed counterpart. When ϕ acquires a VEV, these couplings turn into fermion mass terms proportional to the product of the VEV of the Higgs field and the corresponding Yukawa coupling. The strengths of these couplings are distinct for each fermion, leading to their different masses, with no theoretical backing as to *why* each Yukawa coupling takes the value it does. In other words, while the Higgs mechanism can successfully explain how particle masses arise once the Higgs field obtains a nonzero VEV, it does not offer a deeper principle that predicts these numerical values. Each Yukawa coupling becomes an independent free parameter of the theory, fit to experimental data rather than derived from first principles. This constitutes a key limitation of the SM, which, despite its remarkable success in describing observed phenomena, ultimately leaves unanswered the fundamental question of why the fermion masses span such a wide range [22], leaving the door open for theories that could explain or reduce the number of these free parameters.

2.2. A Quick Introduction to Experimental High Energy Physics (HEP)

While the theoretical framework of the SM provides elegant mathematical descriptions of fundamental particles and their interactions, experimental validation remains essential for advancing HEP. Numerous experimental facilities worldwide contribute to this endeavour, including the Fermi National Accelerator Laboratory (Fermilab) in the United States, KEK in Japan, and DESY in Germany. Neutrino observatories such as Super-Kamiokande in Japan and SNO in Canada have made critical discoveries regarding neutrino oscillations [17, 18], while B-factories like BaBar and Belle have probed Charge-Parity (CP) violation in meson systems [27, 28]. Among these experimental efforts, particle colliders remain at the forefront of investigations into the nature of elementary particles, with the LHC at CERN standing as the world's most powerful accelerator complex [29]. This thesis focuses primarily on physics results from proton-proton (pp) collisions, specifically on data recorded by the CMS detector at the LHC, which will be described in detail in Chapter 3. The subsequent sections of this chapter establish foundational concepts in experimental HEP, including coordinate systems, kinematic variables, and the physics of particle jets, all of which are necessary for the reader to interpret collision data and connect experimental observations to theoretical predictions.

2.2.1. Commonly Used Terminology in Experimental HEP

The following section introduces key concepts and terminology used in experimental HEP, and is necessary to understand the technical content of most analyses performed in the context of the experiments stationed at the LHC. Since these will be used throughout this thesis, they are defined here unambiguously.

- **Transverse Momentum (p_{T})**: The component of a particle's momentum perpendicular to the beam axis (by convention, the z direction):

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} \tag{2.1}$$

  Transverse variables play a fundamental role in hadron collider physics for several reasons. First, the exact longitudinal momentum fractions carried by the interacting partons in the initial state are unknown, but the initial transverse momentum is approximately zero. Conservation of momentum in the transverse plane thus provides a powerful constraint for event reconstruction and analysis. Second, transverse quantities remain invariant under Lorentz boosts along the beam direction, making them robust observables that are independent of the center-of-mass frame of the parton-parton collision. Third, significant deviations from momentum conservation in the transverse plane (Missing Transverse Energy (MET)) serve as signatures for weakly interacting particles that escape detection, such as neutrinos or potential BSM particles. Finally, the transverse plane is where the detector has its most comprehensive coverage, allowing for more precise measurements compared to the longitudinal direction where detector acceptance is limited.

- **Rapidity (y)**: Another measure describing the particle's trajectory relative to the beam axis:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \tag{2.2}$$

For massless particles, the approximation $E \approx |\vec{p}|$ holds, and the rapidity reduces to the more-commonly encountered pseudorapidity ($y \approx \eta$) which is defined below.

- **Pseudorapidity (η)**: A spatial coordinate describing the angle of a particle's trajectory relative to the beam axis:
$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{2.3}$$
where θ is the polar angle from the beam axis. Pseudorapidity provides a Lorentz-invariant way to describe angular distributions and can be directly measured by, for example, the CMS tracker.

- **Azimuthal Angle (ϕ)**: The angle in the transverse plane as measured from the x-axis:
$$\phi = \arctan\left(\frac{p_y}{p_x}\right) \tag{2.4}$$
The kinematics of a particle can be described by its momentum four-vector (p_x, p_y, p_z, E) or equivalently, by $(p_\mathrm{T}, \eta, \phi, m)$

- **MET**: The negative vector sum of all visible transverse momenta in an event:
$$\vec{p}_\mathrm{T}^{\,\mathrm{miss}} = -\sum_i \vec{p}_{\mathrm{T,i}} \tag{2.5}$$
This quantity indicates the presence of particles that do not interact with the detector medium, such as neutrinos or potential signatures of new physics.

- **Luminosity ($\mathcal{L}$)**: This quantity is a measure of the collision rate per unit area. The instantaneous luminosity for two Gaussian beams can be defined as:
$$\mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y} \quad [\mathrm{cm}^{-2}\mathrm{s}^{-1}] \tag{2.6}$$
where N_1, N_2 are the number of protons per bunch, f is the revolution frequency, N_b the number of bunches, and σ_x, σ_y the transverse beam sizes. An additional term, the *integrated luminosity*:
$$L_\mathrm{int} = \int \mathcal{L}(t)dt \quad [\mathrm{fb}^{-1}] \tag{2.7}$$
determines the total event count as $N = \sigma \cdot L_\mathrm{int}$ where σ is the cross section of the process of interest. The concept of luminosity is described in detail in Ref. [30].

  In general, the luminosity can be increased by decreasing the beam size, which is achieved by focusing the beam more intensely using quadropole magnets.

- **Jet**: A collimated spray of hadrons originating from a quark or gluon. Due to the phenomenon of colour confinement in QCD, quarks and gluons cannot exist as free particles [20, 21]. Instead, as they move away from the interaction point, they undergo a process called hadronisation where the colour field energy creates quark-antiquark pairs that combine into colourless hadrons. [31, 32]. This results in the characteristic cone-shaped jet structure observed in detectors such as the CMS or ATLAS detectors. Jets are reconstructed using clustering algorithms with a distance parameter typically defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. Jets and their clustering algorithms are further elaborated upon in Section 2.3.

These kinematic quantities form the foundation for physical observables in LHC analyses.

2.3. Jet Physics

In high-energy pp collisions at the LHC, quarks and gluons produced in hard scatterings are not observed directly due to colour confinement. Instead, they manifest as collimated sprays of hadrons or *jets*, as defined in the preceding section. The process that leads from the initial partons to the final-state hadrons involves several stages that are essential for correctly constructing and interpreting jet-related observables in the CMS detector.

The first stage is the emission of additional quarks and gluons through QCD radiation. This part of the process is described by the *parton shower*, a series of successive branchings governed by perturbative QCD and encoded in the DGLAP evolution equations [33]. Because gluons themselves carry colour charge, they can radiate further, creating a cascade of increasingly lower-energy partons.

As the energy scale drops and the virtualities of partons approach $\mathcal{O}(1\ \text{GeV})$, the strong coupling α_s becomes large and perturbative techniques are no longer valid. At this stage, *hadronisation* takes over—the non-perturbative process by which coloured partons are transformed into colour-neutral hadrons. This process is not derived from first principles, but instead modelled phenomenologically.

Two main hadronisation models are widely used : the *Lund String* model treats the colour field between separating partons as a string-like object, which breaks via quark–antiquark pair creation when stretched [31]. In contrast, the *cluster model* groups colour-connected partons into low-mass colour-singlet clusters, which then decay into hadrons [32]. These models are implemented in Monte Carlo event generators such as Pythia and Herwig, which are used to generate simulated datasets in CMS analyses.

Understanding these stages, starting from the perturbative parton shower to the non-perturbative hadronisation, is crucial for accurate jet reconstruction and for comparing experimental results to theoretical predictions.

2.3.1. Jet Clustering Algorithms

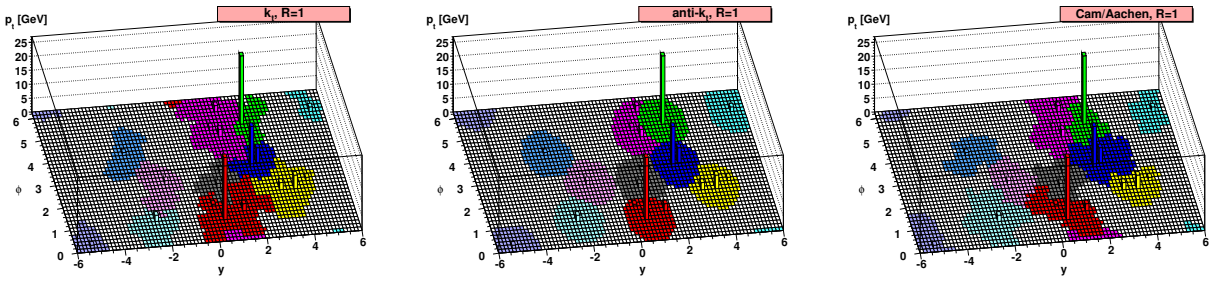
The hadrons resulting from the hadronisation process are not used directly for jet reconstruction in CMS. Instead, jets are built from Particle Flow Candidates (PFCands), which are reconstructed using the Particle Flow (PF) algorithm (described in Section 3.9). To group these constituents into jets, clustering algorithms are employed that follow well-defined, infrared- and collinear-safe procedures.

The most commonly used family of jet algorithms are the sequential recombination algorithms, which operate by iteratively combining particle pairs based on a distance measure defined in the rapidity–azimuth $(y - \phi)$ plane. The generic form of this distance metric is given by [34, 35]:

$$d_{ij} = \min\left(p_{\mathrm{T},i}^{2p},\ p_{\mathrm{T},j}^{2p}\right)\ \frac{\Delta R_{ij}^2}{R^2}, \qquad d_{iB} = p_{\mathrm{T},i}^{2p}, \tag{2.8}$$

where $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ is the squared angular distance between particles i and j, R is the jet radius parameter, and p is a parameter that controls the behaviour of the algorithm. At each iteration, the algorithm identifies the smallest distance, this being either between

two particles or between a particle and the beam, and accordingly merges the closest pair or designates a particle as a final jet.



(a) k_T algorithm **(b)** anti-k_T algorithm **(c)** Cambridge-Aachen algorithm

Figure 2.2.: Examples of jets clustered with 3 different clustering algorithms and radius parameter $R = 1.0$, taken from Ref. [36].

Different choices of the parameter p result in different algorithms with distinct clustering behaviour, as can be seen in Figure 2.2:

- **k_T algorithm ($p = 1$)**: Soft (low p_T) particles are clustered first, which causes soft radiation to be merged before hard substructures. As a result, k_T jets have irregular shapes and boundaries that closely follow the distribution of soft radiation. This behaviour makes the k_T algorithm especially useful in contexts such as pile-up mitigation or grooming studies, where resolving soft components of a jet is important [34, 35].
- **Cambridge/Aachen (C/A) algorithm ($p = 0$)**: Recombination is initiated on the basis of angular proximity, independent of particle p_T. This feature makes C/A jets well-suited for jet substructure analyses. Since the clustering history is determined only by angles, it preserves information about the angular structure of the jet, which is valuable when identifying features like two-prong decays from boosted heavy particles [37, 38].
- **Anti-k_T algorithm ($p = -1$)**: Soft (low p_T) particles are effectively clustered around hard cores, which results in jets with nearly circular, conical shapes. The anti-k_T algorithm has become the default choice in most CMS analyses due to its robustness and the ease with which its jets can be calibrated and corrected [39]. Its uniform jet shapes are particularly advantageous for measurements of inclusive jet spectra and for pile-up subtraction techniques based on jet areas.

Infrared and collinear (IRC) safety: In any realistic collider environment, jet clustering algorithms must be *infrared and collinear safe*. This means the jet definition should remain unchanged if:

- a very soft (low-p_T) particle is added to the event (infrared safety), or
- a particle splits into two collinear particles (collinear safety).

These properties are essential for ensuring that jet observables can be reliably calculated using perturbative QCD, without divergences in the fixed-order cross sections [40, 41].

The distance measure defined in Eq. (2.8) ensures IRC safety across all values of p used in the standard algorithms. In particular, the anti-k_T algorithm has an additional feature of *soft resilience*, meaning that soft particles do not distort the shape or direction of the jet; instead, they are simply incorporated into the area of the nearest hard jet [39].

Practical considerations: Efficient implementations of these algorithms are nowadays available through the FastJet package, which provides an $O(N \log N)$ clustering speed, in addition to tools for computing jet areas and pile-up corrections [42]. In standard pp collisions with high pile-up, anti-k_T jets with a radius parameter $R = 0.4$ (AK4 jets) are the preferred choice, as they capture most of the jet's energy while simultaneously excluding the soft underlying event and noise from pile-up.

In heavy-ion collisions or low p_T environments, larger radius parameters and algorithms like k_T are preferred to better capture diffuse energy deposition. The C/A algorithm is often employed when the internal structure of the jet, in particular the presence of subjets, is of primary interest, especially in searches for boosted objects or jet tagging [43].

Choosing the appropriate clustering algorithm and radius parameter is therefore context-dependent and guided by the physics goals of the analysis, the event topology, and the characteristics of the background environment.

The concepts introduced and elaborated in this chapter are fundamental to follow the remainder of this thesis. A clear understanding of these theoretical and experimental frameworks, along with the terminologies, methodologies, and physical concepts addressed here provide the necessary context to the reader, ensuring that the motivations and implications of later discussions remain accessible and understandable.

3. The CMS Experiment at CERN

3.1. The Experimental Particle Physics Program at the LHC

The LHC at CERN is, as of 2025, the world's leading collider for HEP, housed in a circular underground tunnel spanning 27.6 km in circumference. During the current phase of its operation, the so-called Run 3 (2022–2026), the LHC accelerates protons to energies of 6.8 TeV per beam, achieving a center-of-mass collision energy of 13.6 TeV. The proton beams are comprised of bunches, each containing approximately $\sim 10^{11}$ particles, which collide at four interaction points where the ATLAS [44], CMS [45], ALICE [46], and LHCb [47] experiments are positioned. Guided by superconducting magnets, collisions occur about every 25 ns, resulting in a raw data output rate of approximately $\sim 40\,\mathrm{PB\,s^{-1}}$, which exceeds the permanent storage capabilities housed at CERN and associated institutions.

To address this, the experiments employ real-time trigger systems that selectively retain $< 0.001\%$ of collisions. The CMS and ATLAS detectors use a two-stage trigger architecture: a hardware-based Level 1 (L1) system processes coarse detector data at 40 MHz, discarding $> 99\%$ of events within microseconds using criteria like the presence of high-p_{T} leptons, jets, or large MET. The events which survive the trigger criteria are then analysed by software-based High-Level Triggers (HLTs), which apply specially designed algorithms to isolate comparatively rarer processes (such as Higgs boson decays) while reducing the output rate to a more-manageable value of about 1 kHz [48, 49].

The LHC currently operates at a peak instantaneous luminosity of $2.2 \times 10^{34}\,\mathrm{cm^{-2}s^{-1}}$, on course to deliver an integrated luminosity of $300\,\mathrm{fb^{-1}}$ by 2025 [29] as part of the ongoing Run 3, surpassing the $\sim 138\,\mathrm{fb^{-1}}$ achieved in Run 2 (2016-2018). This unprecedented collision rate enables precision measurements of SM parameters and searches for BSM physics, underpinned by detector systems capable of resolving sub-nanosecond timing and micrometer-scale spatial precision.

3.1.1. The CMS and ATLAS Collaborations

The CMS and ATLAS detectors are the LHC's two general-purpose experiments, designed to address a broad spectrum of physics objectives, ranging from precision SM measurements to searches for exotic phenomena. While both detectors share core design components for tracking, calorimetry, and muon detection, their technical designs are distinct and complementary.

- The **CMS** Detector employs a compact cylindrical geometry centered around a high-field superconducting solenoid magnet (3.8 T), optimised for high-momentum resolution in the inner tracking system. Its silicon-based tracker and lead tungstate electromagnetic calorimeter prioritize hermeticity and fast response, while the muon system, embedded

in a steel return yoke, combines gas ionisation detectors with resistive plate chambers for efficient triggering. The detector's homogeneous calorimetry and strong magnetic field make it particularly suited for precise energy measurement and MET reconstruction.

- The **ATLAS** Detector, on the other hand, features a larger-radius toroidal magnet system, enabling standalone muon momentum measurement with minimal material interference. Its liquid argon calorimeters provide fine granularity and stability over a wide pseudorapidity range, complemented by a tile-based hadronic calorimeter in the central region. The inner tracker combines pixel and strip detectors for robust pattern recognition, while the muon spectrometer leverages monitored drift tubes and resistive chambers for high-precision tracking. This configuration favors long-lived particle searches and forward physics.

After the discovery of the Higgs boson in 2012 as mentioned earlier, subsequent analyses by the ATLAS and CMS Collaborations have achieved precision measurements of the Higgs boson's mass, spin-parity properties, and couplings to fermions and gauge bosons [50].

A full technical description of the CMS detector is provided in this chapter. The remainder of this thesis focuses on and utilises the standard terminology of the CMS Collaboration, as the work described in it was performed as part of the CMS Analysis **CMS-EXO-22-026** [51].

3.2. Introduction to the CMS Experiment

Among the four major experiments stationed at the LHC, the CMS experiment [52] is a general-purpose experiment, centred around the CMS detector designed to investigate a broad spectrum of physics processes. Its primary goals include studying the properties of the Higgs boson, exploring physics beyond the SM, and performing precision measurements of the SM, offering extensive opportunities for scientific discovery.

Collisions at the LHC occur every 25 ns, resulting in an interaction frequency of 40 MHz. At each collision point, two opposing proton bunches interact, generating multiple interactions simultaneously, with the softer collisions usually referred to as *pile-up*. The actual collision itself is a high-energy interaction where protons break apart and their constituent partons interact, creating new particles. Each such recorded collision is therefore called an *event*. Subject to passing specific trigger criteria, these events are recorded, timestamped, and stored for further detailed offline analysis.

The core feature of the CMS detector is its powerful superconducting solenoid magnet, providing a uniform magnetic field of 3.8 T. This magnet is a cylindrical coil with an internal diameter of approximately 6 m and a length of 13 m, making it the largest superconducting magnet ever constructed for a particle physics experiment. This intense magnetic field bends the trajectories of charged particles, allowing precise measurements of their transverse momentum.

Embedded within and surrounding this solenoid magnet are various detector subsystems that enable comprehensive measurement and identification of particles. These include highly sensitive tracking detectors to measure charged particle trajectories, electromagnetic and hadronic calorimeters for energy measurements of photons, electrons, and hadrons, and an extensive muon detection system built within the return yoke of the magnet. Together, these components ensure nearly complete angular coverage, allowing the CMS experiment to detect

and precisely measure the properties of particles produced in collisions, thereby advancing our understanding of particle physics.

3.3. Detector Coordinate System

The standard practice within the CMS Collaboration is to employ a right-handed coordinate system. The origin is the nominal interaction point at the centre of the detector. The x-axis points radially towards the centre of the LHC ring, the y-axis points vertically upward, and the z-axis follows the anticlockwise beam direction.[1] Cylindrical coordinates (r, ϕ) are used in the transverse plane, where ϕ denotes the azimuthal angle about the z-axis. The polar angle θ is measured with respect to the $+z$ direction, and pseudorapidity is defined as $\eta = -\ln\tan(\theta/2)$. The mean-squared difference can then be combined into the Lorentz-invariant quantity $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, crucial for jet-clustering and lepton-isolation algorithms.

3.3.1. Tracking

In the context of the CMS detector, *tracking* refers to the process of reconstructing the trajectories of charged particles produced in proton-proton collisions. As charged particles traverse the silicon sensors within the detector, they deposit localized signals known as hits. Tracking algorithms utilise these hits by identifying patterns and fitting them to determine each particle's trajectory. The high precision of the silicon detectors, coupled with the strong magnetic field generated by the superconducting solenoid, enables accurate measurements of the curvature of charged particle paths. This curvature directly provides a precise estimate of the particle's p_{T}, with a resolution better than 1% for isolated muons with momenta around 100 GeV in the central region of the detector [53, 54].

In addition to measuring particle momenta, tracking algorithms also play a crucial role in vertex reconstruction. Vertexing methods use clusters of reconstructed particle trajectories to determine the collision point (primary vertex) and to identify secondary vertices, indicative of displaced decay processes such as those involving bottom (b) quarks or tau (τ) leptons [55]. Precise vertex reconstruction is essential for numerous physics analyses, especially those focused on studying heavy-flavour physics and identifying new particles with displaced decay signatures.

3.3.2. Barrel and Endcap Regions

Given that proton beams at the LHC travel and collide along the z-axis, the design of the CMS detector naturally leads to distinct geometrical regions. The detector is commonly segmented into two principal regions: the *barrel region*, corresponding to the cylindrical volume around the collision point ($|z| \ll r$), and the *endcap regions*, which are located at each end of the cylindrical detector volume, covering forward angles ($|z| \gtrsim r$). The transition from barrel to endcap occurs approximately at pseudorapidity values around $|\eta| \sim 1.2$ to 1.5. At this point, the

[1] This definition follows the standard LHC convention and differs from the Tevatron convention by the direction of the z-axis.

geometry of detector elements shifts from predominantly cylindrical structures in the barrel region to disk-shaped planar structures perpendicular to the beam axis in the endcaps [52].

This segmentation into barrel and endcap regions ensures comprehensive and uniform coverage in particle detection across a broad angular range. It also facilitates efficient handling of the differing particle flux and radiation levels encountered at varying distances and angles from the collision point. All subdetectors within the CMS Detector such as the tracker, calorimeters, and muon systems are designed and optimised with respect to these distinct geometrical regions, ensuring effective detection and reconstruction capabilities across the full detector acceptance [56, 57, 58].

3.4. The Inner Tracking System

The inner tracking system of the CMS detector [53] is designed to precisely reconstruct the trajectories of charged particles produced in proton-proton collisions. Occupying a cylindrical region surrounding the collision point, it is fully enclosed within the superconducting solenoid magnet. This placement allows the tracker to measure the curvature of charged particle trajectories, enabling accurate determination of their momentum.

The tracker system primarily employs silicon sensors, chosen for their fine spatial resolution and ability to withstand high radiation levels near the interaction region. It is segmented into multiple layers and disks arranged in concentric cylindrical and disk-shaped structures, providing comprehensive spatial coverage in both the radial and longitudinal directions.

The tracker is broadly divided into two main subsystems:

- **Pixel Tracker:** The innermost subsystem consists of silicon pixel detectors, characterised by their extremely fine segmentation. The pixel tracker comprises several concentric cylindrical layers (barrel region) close to the interaction point, complemented by disk-shaped pixel layers in the forward regions (endcaps). Its main function is the accurate measurement of particle trajectories close to their point of origin, essential for precise vertex reconstruction and identification of short-lived particles.
- **Strip Tracker:** Surrounding the pixel tracker, the strip tracker uses silicon sensors with elongated segments (strips), which cover larger areas. This subsystem is further divided into inner and outer barrels in the central region, and disk-like structures in the forward regions, known as Tracker Endcaps. Its primary function is to measure particle trajectories at intermediate and larger radii from the interaction point, thus extending the precise determination of particle momenta outwards.

Altogether, the inner tracker has about 200 m^2 of silicon sensors, and $O(10^9)$ read-out channels. This allows for high granularity measurement and tracking of particles, ultimately enabling critical aspects of the CMS physics program such as precision measurements of the SM or searches for rare new physics.

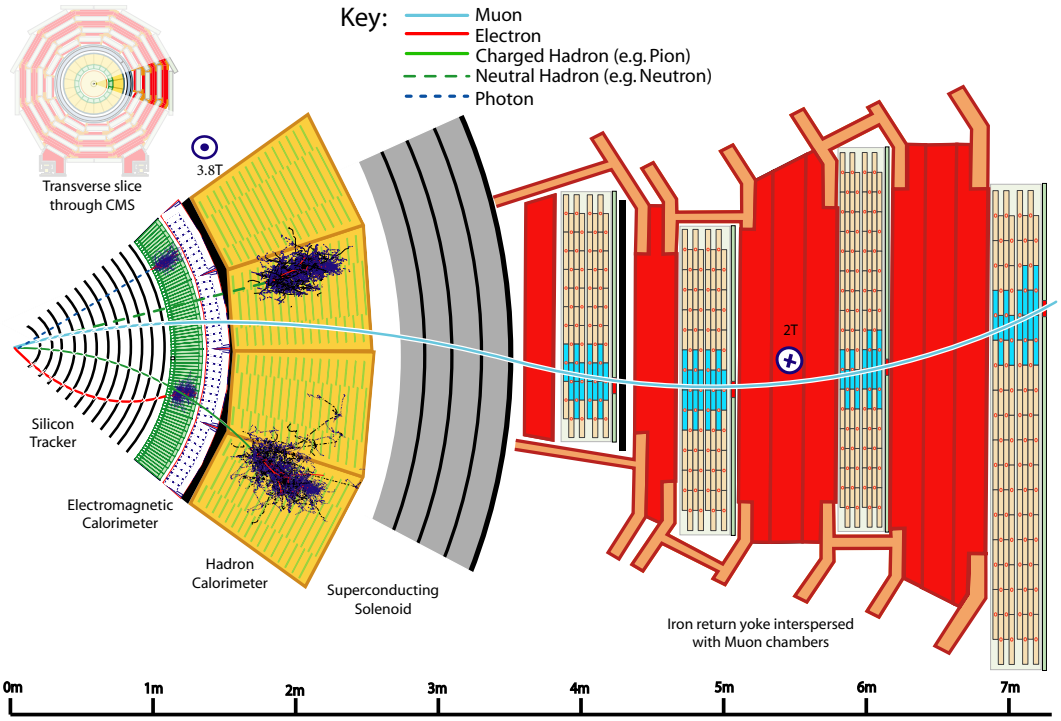


Figure 3.1.: A schematic showing each layer of the CMS detector, along with the various possible particle interactions [59]

3.5. ECAL

The CMS ECAL is a homogeneous scintillating-crystal calorimeter that employs lead-tungstate ($PbWO_4$) crystals as the active medium [56]. It is specifically designed to measure the energies of electrons and photons produced in pp collisions with high precision. The ECAL's energy resolution capability and fine granularity are crucial for reconstructing the decay products of Higgs bosons, in particular the $H \to \gamma\gamma$ and $H \to ZZ^{(*)} \to 4\ell$ channels [5].

Table 3.1.: Principal design parameters of the ECAL.

Quantity	Barrel	Endcap
$\lvert\eta\rvert$ coverage	< 1.479	1.479–3.0
Number of crystals	61 200	14 648
Crystal length	230 mm (25.8 X_0)	220 mm
Front-face area	22×22 mm^2	28.6×28.6 mm^2
Inner radius	129 cm	—

Each $PbWO_4$ crystal emits scintillation light when traversed by high-energy electromagnetic particles. The emitted light is proportional to the deposited energy, and is collected by Avalanche Photo Diodes (APDs) in the barrel and vacuum phototriodes in the endcaps. The combination of the crystals' short radiation length (X_0), small Molière radius, and the finely segmented readout enables precise localisation of electromagnetic showers and excellent energy resolution. A preshower detector is positioned in front of each ECAL endcap, covering the pseudorapidity range $1.653 \leq |\eta| \leq 2.6$. It consists of two layers of silicon strip

sensors interleaved with 2 mm thick lead absorbers, which contribute to improving the π^0/γ discrimination.

3.6. Hadron Calorimeter (HCAL)

The CMS HCAL [57] is responsible for measuring the energy of hadrons, as they are typically produced in abundance in high-energy proton-proton collisions. Unlike electrons or photons that deposit energy primarily in the electromagnetic calorimeter, hadrons penetrate deeper into the detector and require a dense, thick absorber to be fully contained.

The HCAL is a sampling calorimeter, consisting of alternating layers of dense absorber material and active scintillator elements. When hadrons pass through the absorber layers (typically made of brass or steel), they initiate hadronic showers, with cascades of secondary particles. These secondary particles then interact with the scintillator tiles, producing light signals that are collected via wavelength-shifting fibers and read out by photodetectors.

To effectively capture these showers, the HCAL is segmented into different regions based on geometry and detector placement:

Barrel and Endcap HCAL: These regions cover the central and intermediate angular zones of the detector. Positioned just outside the electromagnetic calorimeter and within the solenoid magnet, they provide sufficient material to absorb most hadronic showers in this angular range.

Outer HCAL: Since the solenoid magnet itself is not part of the calorimetry system, it introduces a layer of material that can cause some energy leakage. It is installed outside the magnet coil, embedded within the return yoke, to capture any remaining energy and improve measurement accuracy in the central region.

Forward HCAL: Located far downstream along the beamline, the HF is designed to detect particles emitted at very small angles relative to the beam direction. It uses quartz fibers embedded in a steel matrix to detect light from charged particles moving faster than the speed of light in the medium, producing Cherenkov radiation. It plays a crucial role in ensuring full angular coverage and contributes significantly to the measurement of MET.

Together, these HCAL components ensure a high degree of hermeticity, which is vital for reconstructing the total energy flow in events and for detecting imbalances that may indicate undetected particles, such as neutrinos or possible BSM particles.

3.7. Muon System

The CMS detector was specifically designed to have the capability to detect muons and measure their properties, as reflected in its name: *Compact Muon Solenoid.* Muons are particularly important because they penetrate through all other detector layers with minimal energy loss, reaching the outermost regions of the detector. This property makes them excellent probes for many key physics signatures, including decays of heavy particles like the Z or Higgs bosons, or even hypothetical BSM resonances.

The muon system is embedded in the massive steel return yoke of the solenoid magnet, which serves a dual purpose: guiding magnetic flux and acting as an absorber for other particles, allowing only muons to reach the muon detectors. The system consists of several layers (stations) of gaseous detectors that provide precise position and timing measurements. These allow the muon's trajectory to be reconstructed and matched to tracks found in the inner tracker.

To cope with different rates of particle flux and to provide redundancy, three different detector technologies are used:

- **Drift Tubes (DT):** These are used in the central barrel region where the particle flux is relatively low. They offer high spatial resolution and are well-suited for covering large volumes.
- **Cathode Strip Chambers (CSC):** Located in the endcaps where particle rates and magnetic field inhomogeneities are higher, CSCs are capable of operating under more challenging conditions while providing precise measurements.
- **Resistive Plate Chambers (RPC):** These detectors are installed in both barrel and endcap regions, offering excellent time resolution. They are used primarily for fast triggering and for enhancing the robustness of muon identification.

Recent upgrades in anticipation of the High Luminosity LHC (HL-LHC) have introduced new technologies, such as Gas Electron Multipliers (GEMs), in the forward region. These detectors are designed to withstand the higher radiation and occupancy expected in future LHC runs, and will complement the existing systems by improving coverage and resolution where it is most needed [58].

Together, the muon system provides redundant and robust tracking and identification of muons over a wide range of pseudorapidities. Its high spatial and temporal resolution make it a crucial component for both triggering and precision physics measurements.

3.8. CMS Trigger System

At the LHC, proton bunches cross one another tens of millions of times per second, with each crossing producing multiple interactions. Recording *all* of these crossings would require a continuous data stream of many tens of terabytes per second, well beyond the capabilities of present-day storage or offline processing systems. To cope with this challenge, CMS employs a multi-tiered *trigger* and data-acquisition (TDAQ) architecture that selects, in real time, only the most interesting collisions for permanent storage [60, 61].

Level-1 Trigger (L1): Implemented entirely in custom electronics based on Application Specific Integrated Circuits (ASICs) and Field Programmable Gate Arrays (FPGAs), the L1 trigger receives coarse information from the calorimeters and muon detectors at the full 40 MHz bunch-crossing rate. Within a few micro-seconds, it decides whether an event shows basic signatures of potentially interesting physics, these could for example be highly energetic jets, energetic leptons, or large values of MET. A rate of only about 100 kHz worth of events are accepted for further scrutiny, reducing the data flow by roughly three orders of magnitude.

High-Level Trigger (HLT): Events accepted by the L1 trigger are transferred to a large farm of commodity servers. Here, software algorithms, closely related to those used in offline reconstruction, re-examine each event with finer granularity. By applying more sophisticated selections, the HLT trims the event rate from 100 kHz to about 0.8 kHz to 1 kHz. This final rate is compatible with long-term data storage and with the throughput of the worldwide offline computing grid.

Events that survive both trigger levels are written to CERN's Tier-0 computing centre, buffered on high-speed disks, and then distributed to regional Tier-1 and Tier-2 centres for full reconstruction and analysis. Continuous data-quality monitoring ensures that any anomalies in detector performance or software reconstruction are identified in real time.

3.9. The CMS Particle Flow Algorithm

At the LHC, pp collisions produce a large number of final-state particles, including many originating from overlapping interactions in the same bunch crossing (pile-up). Reconstructing these particles with high precision is essential for all physics analyses. A detector like CMS, with multiple sub-systems optimised for different particle types, requires the combination of information across detectors to build a global picture of the event. This is the core idea behind the PF algorithm [62].

Traditional calorimeter-only reconstruction methods are limited by the relatively poor energy resolution for charged hadrons and by challenges in distinguishing particles from different vertices. In contrast, the CMS tracking system measures the momentum of charged particles with excellent resolution, especially at low to moderate p_{T}, and can associate them to specific primary vertices. The PF algorithm was developed to make full use of the full granularity and redundancy of all CMS sub-detectors, namely the tracker, electromagnetic and hadronic calorimeters, muon chambers, and solenoid, to ultimately reconstruct and identify each stable particle individually, these being the electrons, muons, photons, charged hadrons, and neutral hadrons [63, 64].

First, individual low-level detector signals, such as charged particle tracks from the silicon tracker, energy clusters from the ECAL and HCAL, and segments from the muon system, are collected and promoted to *PF elements*. A dedicated linking algorithm then associates elements, in pairs, that are geometrically and energetically compatible and likely to originate from the same physical particle. For example, a track extrapolated to the calorimeter is matched to clusters in the ECAL and/or HCAL, and muon system segments are matched to tracks in the tracker. These connected sets of elements are grouped into *PF blocks*, which represent localized regions of the detector potentially containing one or more particles.

Within each PF block, a sequence of reconstruction steps is applied to identify and reconstruct individual particles. The algorithm proceeds in a fixed order: it first identifies muons and electrons by combining tracking and calorimetric or muon information, then reconstructs charged hadrons by combining tracks with calorimeter deposits, and finally identifies photons and neutral hadrons using unassociated calorimeter clusters.

After this identification pass, a final postprocessing step is applied to ensure consistency and improve the overall quality of the reconstruction, in particular to reduce events with an

artificially inflated METdue to misidentified or misreconstructed muons. This postprocessing step is crucial for achieving the high precision required in downstream physics analyses.

This particle-level description leads to improved performance in many key observables. For example, by using tracker momentum measurements instead of calorimeter energy for charged hadrons, the jet energy resolution is significantly enhanced. Similarly, METreconstruction benefits from the improved resolution and vertex association. The PF approach also provides a natural basis for advanced pile-up mitigation strategies, such as the PUPPI algorithm [65]. Further down the line, the PFCands reconstructed by the PF Algorithm are used as inputs to the various jet clustering algorithms that were described in Section 2.3.1.

For the future, CMS is exploring ML variants of the PF Algorithm, using graph neural networks [66]. These approaches learn to reconstruct particles directly from detector signals, offering potential gains in performance and scalability for the HL-LHC era [67]. However the basic paradigm, which is to build events from particles rather than detector signals, remains foundational to CMS data reconstruction and analysis.

3.10. Outlook for the Future

To conclude, the CMS detector is a complex, multi-layered apparatus that combines precise silicon tracking systems, highly granular calorimeters for energy measurements, and robust muon detection systems, all enclosed within a powerful superconducting solenoid. This design enables the experiment to identify and measure the properties of particles produced in high-energy pp collisions with exceptional precision and reliability.

Since the first era of data-taking in 2010, the CMS Collaboration has played a pivotal role in some of the most important discoveries in modern physics. Most notably, it contributed to the discovery of the Higgs boson [5] in 2012, a breakthrough that confirmed the last missing piece of the SM and led to the 2013 Nobel Prize in Physics being awarded to François Englert and Peter Higgs. Beyond this landmark achievement, the research activities of the CMS Collaboration have provided crucial insights into the properties of the Higgs boson, precision measurements of the electroweak sector, top quark physics, and searches for new phenomena such as supersymmetry and dark matter candidates [68, 69, 70].

The LHC's experimental program, driven by CMS, ATLAS, and other collaborations, combines cutting-edge detector innovation with high-statistics datasets to probe the SM, and search for BSM physics. The ongoing HL-LHC upgrades are planned to extend its discovery potential into the 2030s. The HL-LHC is designed to deliver an integrated luminosity of up to 3000–4000 fb^{-1} [71, 72], enabling unprecedented precision measurements of SM parameters and greatly increasing sensitivity to rare processes and heavy new particles. To cope with the significantly higher pile-up and radiation conditions, both CMS and ATLAS are undergoing major detector upgrades, including new high-granularity tracking systems, precision timing detectors, and enhanced trigger capabilities [73, 74, 75, 76]. These improvements are essential not only for sustaining performance in the harsh HL-LHC environment, but also for advancing the scope and reach of future analyses, and ultimately cementing the LHC's position at the forefront of particle physics for the next two decades.

With these in mind, the CMS detector is poised to meet the challenges of the HL-LHC era. The increased collision rates and data volumes will push the detector systems to their limits, but a comprehensive program of upgrades, involving new tracking detectors, improved calorimeters, extended muon coverage, and advanced trigger and data acquisition systems is already underway. These efforts aim not only to maintain performance but also to enhance sensitivity to rare signals and deviations from the SM. As a result, the CMS detector remains at the forefront of discovery potential, well-positioned to uncover evidence of BSM physics over the next two decades.

Part III.

Machine Learning for Unsupervised Anomaly Detection at the CMS Experiment

4. A Quick Look at Machine Learning

4.1. Introduction: ML and HEP

Machine Learning (ML) refers to a broad set of computational techniques that enable algorithms to learn patterns and make predictions or decisions directly from data, rather than relying on explicitly programmed rules. At its core, ML seeks to approximate functions or identify structure in high-dimensional spaces by iteratively improving a model's internal parameters based on data-driven feedback. The historical development of ML spans foundational milestones such as the perceptron model [77], the introduction of backpropagation [78], and the success of deep convolutional Neural Networks (NNs) in computer vision [79]. These advances have gradually positioned ML algorithms as powerful tools capable of learning meaningful representations from large, noisy, and complex datasets.

In HEP, and particularly within the context of the CMS experiment at the LHC, ML has rapidly become indispensable across a wide range of applications. As the volume and complexity of collision data grows, traditional analysis methods are increasingly constrained by their reliance on specific features and fixed selection criteria. By contrast, ML models offer flexible, data-driven alternatives that can optimise performance across tasks such as classification, regression, and anomaly detection. For instance, the identification of jets originating from b-quarks now relies heavily on DNNs such as DeepCSV and DeepJet [80]. These networks integrate information from multiple sub-detectors and leverage both low-level and high-level features. Similarly, ML architectures that learn directly from point cloud representations, such as ParticleNet [81] and ParticleTransformer [82], have significantly improved jet tagging performance by learning from raw particle-level information. Graph NNs, which model the topological relationships between detector hits, have shown promise in track reconstruction tasks under the challenging high-pileup conditions expected at the HL-LHC [83]. Beyond reconstruction, unsupervised models such as autoencoders and variational networks have been studied for deployment in online systems for real-time anomaly detection and decision making in triggers [84, 85], enabling the identification of rare signatures with ultra-low latency.

These applications reflect a broader trend across the field: as data volumes increase and detector systems grow more complex, ML is no longer just a useful tool but a foundational component of experimental workflows. Recent reviews [86, 87, 88] have highlighted this, noting that the demands of modern HEP experiments offer a lot of potential, not just for improving sensitivity and efficiency in existing analyses, but also for enabling entirely new approaches to data processing, and interpretation.

4.2. Learning Paradigms in ML

ML encompasses a range of methodologies designed to extract patterns and make predictions from data. Depending on the nature of the available data and the objectives of the analysis, there are a number of learning paradigms to choose from, as summarised below. In HEP, this choice is made when it comes to complex tasks such as jet classification and the search for BSM physics.

- **Supervised Learning:** This paradigm involves training models on datasets where each input has a known label. The supervised model is expected to learn this map between inputs and the corresponding labels, thereby enabling it to predict the corresponding labels for new, unseen data. In HEP, supervised learning has been effectively applied to jet classification tasks, where the goal is to distinguish between jets originating from different particles, the simplest example being from quarks and gluons. For instance, convolutional NNs have been trained on simulated data to classify jet images, achieving high accuracy in identifying jet substructures [89].
- **Unsupervised Learning:** This paradigm is designed to work with data that does not have any known labels. The goal is then to make the model learn inherent structures or patterns within the data. This approach is crucial in anomaly detection, where the objective is to identify events that deviate from the norm (which for particle physicists is simply the SM), potentially indicating new physics phenomena. Techniques such as autoencoders (which will be elaborated on in Section 4.6) and clustering algorithms have been employed to detect anomalies in collision data without relying on predefined labels [90].
- **Weakly Supervised Learning:** This paradigm operates in the area between supervised and unsupervised learning. It utilises datasets with incomplete or incorrect labels. In HEP, weakly supervised learning is advantageous when dealing with datasets that have ambiguous truth labels. This approach has been used in the past at the LHC to enhance sensitivity to new physics in dijet resonance searches [91].

Each of these paradigms offers unique advantages and is chosen based on the specific requirements and constraints of the HEP analysis at hand. The integration of these ML techniques is what enables more precise measurements and the potential discovery of new physical phenomena.

4.3. Fundamental Concepts and Terminology in Machine Learning

Throughout this thesis, various terms and concepts from ML will be employed to describe methodologies and analyses. To ensure clarity and consistency, this section provides definitions and explanations of key ML terminology relevant to the discussions that follow.

- **Model:** A mathematical function that maps input data to output predictions. In supervised learning, models are trained to approximate the relationship between inputs and known outputs.

- **Parameters:** Variables of the model that are learned from data during the course of training. For example, in NN, these are the weights and biases that are adjusted to minimise the loss function.
- **Hyperparameters:** Configuration parameters that are fixed before training begins and are not learned from the data. Examples include the learning rate, batch size, number of epochs, and the architecture of a NN. Hyperparameter tuning is often necessary to obtain the best possible model performance.
- **Loss Function:** A function that quantifies the difference between what the model predicts, versus what the target goal is. Common loss functions include mean squared error for regression tasks and cross-entropy for classification tasks.
- **Optimisation Algorithm:** A method used to adjust the model's parameters to minimise the loss function. Gradient descent and its variants, such as stochastic gradient descent (SGD) along with more advanced algorithms like Adam [92], are commonly used.
- **Learning Rate:** A hyperparameter that determines the step size at each iteration while moving toward the loss minimum. A suitable learning rate is crucial for efficient training, because too high a value can cause divergence or oscillation about the minimum, whereas too low a value leads to slow convergence.
- **Batch Size:** The number of training examples utilised in one iteration of the optimisation process. One may use the entire dataset (batch), a single example (stochastic), or subsets of the data (mini-batch), with the latter being commonly used as it strikes a balance between efficiency and stability.
- **Epoch:** One complete pass through the entire training dataset followed by a validation pass. Training a model involves multiple epochs, allowing the model to learn from the data iteratively.
- **Iteration:** A single update of the model's parameters. The number of iterations per epoch depends therefore on the batch size and the total number of training examples.
- **Overfitting:** A modeling error that occurs when a highly complex model with a large number of parameters *over-learns* the training data, and memorises its artefacts such as noise and outliers, leading to poor performance on unseen data.
- **Underfitting:** A modeling error that occurs when a model is too simple and does not have enough free parameters to adequately learn the underlying representation, resulting in poor performance on both training and test data.
- **Generalisation:** The ability of a model to perform well on new, unseen data. A model with good generalisation captures the underlying patterns in the data without overfitting.
- **Regularisation:** Methods used to prevent overfitting by adding additional information or constraints to the model. Common regularisation methods include L1 and L2 regularisation, dropout, and early stopping. These improve the model's generalisation capabilities when faced with data it has not seen before.
- **Bias-Variance Trade-off:** The balance between the error introduced by approximating a real-world problem (bias) and the error introduced by sensitivity to fluctuations in the training set (variance). Managing this trade-off is crucial for building models that generalise well.

- **Feature:** Each individual property or characteristic of the data. Features are the inputs for the model.
- **Label:** The output or target variable that the model should learn to predict correctly. In supervised learning, each input example is associated with a label.
- **Training Set:** A subset of the dataset used to train the model.
- **Validation Set:** A subset of the dataset used to gauge performance after each epoch and tune hyperparameters when necessary.
- **Test Set:** A subset of the dataset used to assess the final performance of the trained model on unseen data.
- **Early Stopping:** A regularisation technique where training is halted when the model's performance on the validation set stops improving, preventing overfitting.
- **Dropout:** A regularisation technique where a specific fraction of neurons (which in itself is a hyperparameter) is randomly selected and their outputs ignored during training, reducing the risk of overfitting.
- **Momentum:** An optimisation technique that speeds up gradient descent by adding a fraction of the previous update value to the current, smoothening the optimisation path and potentially leading to faster convergence.
- **Activation Function:** A function applied to the output of each neuron, essential for making the output of the network non-linear. Common activation functions include the Rectified Linear Unit (ReLU), sigmoid, and tanh.
- **Backpropagation:** An algorithm used to compute the gradient of the loss function with respect to each weight in the network, enabling efficient computation of gradients for optimisation.
- **ROC Curve and AUC:** A curve used to evaluate the performance of classification models by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various thresholds, with AUC simply being the area underneath it.

These terms are necessary to follow the description of the methodologies, and the results presented in this thesis.

4.4. An Introduction to the DNN

There exist various categories of ML models such as decision trees, support vector machines, ensemble methods like XGBoost, and finally the NN. Each model type has its own strengths and is suited to specific types of problems. In this thesis, the focus is on deep learning models, particularly the DNN, due to its capacity to model complex, high-dimensional data, a common characteristic of HEP datasets.

4.4.1. Historical Context and Development

The conceptual foundation of NNs began with the biological neuron. Thereafter came the perceptron, which was capable of learning binary classifiers through supervised learning [77]. However, the perceptron was limited to solving linearly separable problems. This limitation was addressed by the introduction of the Multi-Layer Perceptron (MLP), which incorporated multiple hidden layers and nonlinear activation functions, enabling the modeling of more complex, non-linear relationships.

4.4.2. Formal Definition of a Neural Network

An NN can be viewed as a parameterised function that maps input vectors to output targets. The core building block of an NN is the neuron, which takes an input $\mathbf{x} \in \mathbb{R}^n$ and computes a weighted sum of this input plus a bias term. The output is then passed through a non-linear activation function ϕ. For a single neuron indexed by j:

$$y_j = \phi\left(\sum_{j=1}^{n} w_{ji} x_i + b_j\right) \tag{4.1}$$

where $w_{ji} \in \mathbb{R}$ is the weight associated with the connection from input i (which could itself be the output of a neuron in the previous layer) to neuron j, and $b_j \in \mathbb{R}$ is the bias of the j^{th} neuron.

One can then set up a layer of m neurons, each performing the above computation independently. The inputs $\mathbf{x} \in \mathbb{R}^n$, which for example could be the output of a layer with n neurons, are mapped to an output vector $\mathbf{y} \in \mathbb{R}^m$ whose j^{th} component is then given by what is already defined in Equation 4.1. Collectively, this can be written in vectorised form as

$$\mathbf{y} = \phi\left(W\mathbf{x} + \mathbf{b}\right), \tag{4.2}$$

where $W \in \mathbb{R}^{m \times n}$ has components w_{ji}, and ϕ is applied element-wise to the vector.

With the above recipe in hand, a DNN is formed by stacking multiple such layers. Let the input be $\mathbf{h}^{(0)} = \mathbf{x}$, and for each layer $\ell = 1, \ldots, L$, the transformation is defined as

$$h_j^{(\ell)} = \phi^{(\ell)}\left(\sum_{i=1}^{n_{\ell-1}} w_{ji}^{(\ell)} h_i^{(\ell-1)} + b_j^{(\ell)}\right), \quad j = 1, \ldots, n_\ell, \tag{4.3}$$

where n_ℓ is the number of neurons in layer ℓ, and $\left(w_{ji}^{(\ell)}, b_j^{(\ell)}\right)$ are the weights and biases for layer ℓ. In vector notation, this is:

$$\mathbf{h}^{(\ell)} = \phi^{(\ell)}\left(W^{(\ell)}\mathbf{h}^{(\ell-1)} + \mathbf{b}^{(\ell)}\right). \tag{4.4}$$

The output of the network is then given by $\mathbf{h}^{(L)} = f(\mathbf{x}; \boldsymbol{\theta})$, where $\boldsymbol{\theta} = \left\{W^{(\ell)}, \mathbf{b}^{(\ell)}\right\}_{\ell=1}^{L}$ is the set of all trainable parameters and f is the learned representation.

An NN is trained by minimising a well-defined loss function $\mathcal{L}(f(\mathbf{x}; \boldsymbol{\theta}), \mathbf{y})$ that quantifies the discrepancy between the predicted output and the desired target $\mathbf{y}$. The loss function depends

on the class of problem that the user wishes to solve. As an example, for the task of q/g jet classification, where the desired outputs are labels $y_{\text{pred}} \in \{0, 1\}$, the final layer would ideally comprise a single neuron, with an activation function bounded in $(0, 1)$ such as the sigmoid (Equation 4.5), along with an appropriate loss function such as the Binary Cross Entropy (BCE) (Equation 4.6) that maximally penalises mislabelling.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{4.5}$$

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^{N} \left[y_{\text{true}} \log(y_{\text{pred}}) + (1 - y_{\text{true}}) \log(1 - y_{\text{pred}}) \right] \tag{4.6}$$

The end goal of training is to find the best parameters $\boldsymbol{\theta}$ that minimise the loss as follows:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}} \left[\mathcal{L}(f(\mathbf{x}; \boldsymbol{\theta}), \mathbf{y}) \right] . \tag{4.7}$$

This is typically done using an optimisation algorithm such as Stochastic Gradient Descent (SGD) [93] or Adam [92]. The gradients $\partial\mathcal{L}/\partial\boldsymbol{\theta}$ are computed using backpropagation [78], which is described in A.1 in detail.

The text in the preceding sections should be sufficient and self-contained to the extent that a reader can understand the discussions centred around ML in the following chapters. It is now necessary to understand the architectures required to perform effective anomaly detection in the context of HEP.

4.5. Model Agnostic Searches

While many searches for BSM physics are guided by specific theoretical frameworks, such targeted approaches risk overlooking unanticipated signatures if new physics manifests in an unexpected way [94]. Model-agnostic machine learning (ML) techniques are therefore becoming increasingly vital. By aiming to identify deviations from the well-established SM without *a priori* assumptions about the nature of new phenomena, these methods cast a wider net, significantly enhancing the potential for discovering novel physics in an unbiased manner. This is crucial at the LHC, where the primary goal is to explore the energy frontier and uncover the unknown, rather than solely confirming or refuting existing theoretical proposals.

The necessity for algorithms that operate independently of specific signal models is primarily due to the speculative nature of BSM theories, since the *correct* model describing new physics, if it exists at energies accessible by modern colliders, is unknown. Relying solely on searches optimised for particular models would mean that signatures not conforming to these preconceived notions could be missed, leading to blind spots in the experimental program. Model-agnostic ML, in particular anomaly detection algorithms, offer a powerful alternative by learning directly from the data [95]. These algorithms can be trained, often in an unsupervised manner, on background (SM) data to learn a representation of the background. Consequently, they can flag rare events that exhibit characteristics inconsistent with the SM, potentially pointing to new physics even without previous knowledge of its specific form without any

reliance on a highly specific particular signal hypothesis. Several pioneering searches at the LHC have demonstrated the feasibility and promise of this approach, for instance, using deep autoencoders to identify anomalous patterns in hadronic final states [96, 97]. The broader field of anomaly detection, with its successful applications in areas such as fraud detection, industrial monitoring, and medical diagnostics [98], further highlights the robustness and adaptability of such data-driven discovery paradigms for navigating complex datasets where the signals of interest are unknown by definition. In the following section, two relevant algorithms: the Autoencoder (AE) and the VAE are discussed.

4.6. Autoencoders for Anomaly Detection

Autoencoders represent a powerful class of NN architectures that have been used significantly in anomaly detection tasks, including searches for BSM physics [97]. Unlike supervised learning methods that require labeled examples of anomalies, autoencoders are unsupervised, making them particularly valuable when anomalous patterns are simply unknown, or not unambiguously defined.

4.6.1. Autoencoder Architecture

An autoencoder is a NN designed to learn efficient data encodings by attempting to reconstruct its own inputs [99]. It has two primary components: an encoder that learns a lower-dimensional representation of the initial input, thereby preserving only the most important information, and a decoder that tries to recreate the original input using this compressed lower-dimensional representation, commonly known as the autoencoder *latent space*. The information bottleneck ensures that the network as a whole learns the underlying data distribution(s). The architecture can be represented as:

$$h(\mathbf{z}) = f_\phi(\mathbf{x}) \quad \text{(encoder)} \tag{4.8}$$

$$\hat{\mathbf{x}} = g_\theta\left(h(\mathbf{z})\right) \quad \text{(decoder)} \tag{4.9}$$

where $\mathbf{x}$ represents the input data (such as QCD jet features, if the task is the detection of anomalous jets), $\mathbf{z}$ is the compressed representation in the latent space, and $\hat{\mathbf{x}}$ is the reconstructed output. The functions f_ϕ and g_θ are parameterised by NNs with weights ϕ and θ, respectively.

The critical feature of an autoencoder is its bottleneck, in which the latent space $\mathbf{z}$ typically has a much lower dimensionality than the input space, forcing the network to learn an efficient encoding of the data's essential features [99]. This property makes autoencoders conceptually related to nonlinear generalisations of principal component analysis [100, 101].

4.6.2. Learning Process and Loss Functions

The autoencoder is trained to minimise the reconstruction error between the input $\mathbf{x}$ and its reconstruction $\hat{\mathbf{x}}$. The most common loss function for continuous data is the mean squared error (MSE):

$$\mathcal{L}_{\text{MSE}}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2 \tag{4.10}$$

where n is the dimensionality of the input. For binary or normalised data, the binary cross-entropy loss introduced in Equation 4.6 is often preferred. Thereafter, the network parameters ϕ and θ are optimised using gradient-based methods, typically stochastic gradient descent or its variants like Adam [92]. The training process requires one to find the set of parameters that minimise the chosen loss function as:

$$\phi^*, \theta^* = \arg\min_{\phi,\theta} \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}} \left[\mathcal{L}(\mathbf{x}, g_\theta(f_\phi(\mathbf{x}))) \right] \tag{4.11}$$

where p_{data} represents the data distribution, typically the SM background processes in the context of BSM searches. It must be noted here that unlike maximum likelihood estimation, where one is required to exhaustively minimise the negative log-likelihood to identify the best-fit parameters, this exhaustive minimisation is typically undesirable in a machine learning context. Indeed, fully minimising the loss function often results in overfitting, leading the model to exhibit poor generalisation to new data.

4.6.3. Anomaly Detection with Autoencoders

When applied to anomaly detection, autoencoders exploit a fundamental but simple principle: they learn to reconstruct data similar to what they were trained on while performing poorly on outliers or anomalies [98]. In the context of HEP, an autoencoder is trained exclusively on simulated SM events, learning efficient representations of known physics processes [96]. After training, the reconstruction error serves as the simplest anomaly score:

$$\text{Anomaly Score}(\mathbf{x}) = \mathcal{L}(\mathbf{x}, g_\theta(f_\phi(\mathbf{x}))) \tag{4.12}$$

Events with high reconstruction errors are flagged as potential anomalies, possibly indicating BSM physics. This approach has been successfully implemented in a number of analyses at the LHC, showing promising sensitivity to a variety of hypothetical new physics scenarios [97, 96].

4.6.4. Limitations and Motivation for Variational Autoencoders

Despite their effectiveness, conventional autoencoders exhibit several limitations that can impact their performance in BSM searches:

1. **Discontinuous latent space**: The biggest disadvantage of an autoencoder is the discontinuous and unstructured latent space, which severely limits its ability to learn a true generalisation of the training dataset.
2. **Overfitting**: Without proper regularisation, autoencoders may learn to reproduce noise or specific features of the training data rather than meaningful representations, though this is a problem not unique to autoencoders alone.
3. **Deterministic encoding**: The mapping between input data and latent representation is deterministic and not probabilistic, which may not adequately capture uncertainty or variability in the data [102, 103].

These limitations motivated the development of VAEs, which introduce a probabilistic framework to the autoencoder architecture [102] at the latent space $h(\mathbf{z})$. Rather than encoding inputs as single points in the latent space, VAEs encode them as probability distributions, typically constrained to be Gaussian by using, for example, a Kullback-Leibler Divergence [104] (also see Section 4.7.2). This probabilistic approach offers several advantages: it regularises the latent space, promotes continuity and smoothness, enables interpolation between data points, and can even be used for the generation of new samples, by sampling from the probability distribution that defines the latent space. [103, 102].

4.7. Variational Autoencoders: A Quick Introduction

The VAE establishes a connection to Bayesian inference and information theory, providing a theoretical justification for the probabilistic representation that it learns. These properties make VAEs promising for model-agnostic BSM searches, where the goal is not only to detect anomalies but also to extract physical insights from the identified deviations from SM expectations. The first part of this thesis therefore introduces a novel technique based on a VAE coupled with a DNN based Quantile Regression (QR) for identifying anomalous signatures in dijet events recorded with the CMS Detector at the LHC during the period 2016-18.

4.7.1. Mathematical Framework of VAEs

Unlike conventional autoencoders that map inputs to deterministic points in the latent space, VAEs encode inputs as probability distributions in the latent space, typically Gaussian as shown in the schematic of Figure 4.1. This fundamentally changes how the model represents data and thereafter generates reconstructions.

For an input data point $\mathbf{x}$ (which could be the particle features from a jet), the encoder of a VAE outputs the parameters describing a probability distribution $q_\phi(\mathbf{z}|\mathbf{x})$ over the latent space $\mathbf{z}$, instead of a direct numerical encoding. In the standard implementation, this distribution would be a multivariate Gaussian with a diagonal covariance matrix:

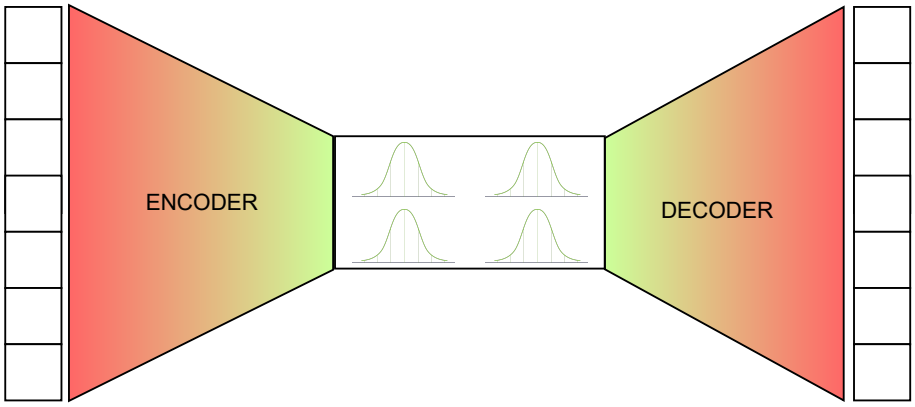


Figure 4.1.: A VAE schematic showing the encoder, Gaussian latent manifold and the decoder

$$q_\phi(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_\phi(\mathbf{x}), \boldsymbol{\sigma}^2_\phi(\mathbf{x})\mathbf{I}) \tag{4.13}$$

$$= \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\sigma}^2_\phi(\mathbf{x})\mathbf{I}|}} \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu}_\phi(\mathbf{x}))^T (\boldsymbol{\sigma}^2_\phi(\mathbf{x})\mathbf{I})^{-1}(\mathbf{z} - \boldsymbol{\mu}_\phi(\mathbf{x}))\right) \tag{4.14}$$

where $\boldsymbol{\mu}_\phi(\mathbf{x})$ and $\boldsymbol{\sigma}^2_\phi(\mathbf{x})$ are the outputs of NNs parameterised by ϕ, representing the mean and variance of the latent distribution for input $\mathbf{x}$, and d is the dimensionality of the latent space.

The decoder then reconstructs the input by mapping samples from this latent distribution back to the original space. It defines another distribution $p_\theta(\mathbf{x}|\mathbf{z})$ over possible reconstructions given the latent vector $\mathbf{z}$:

$$p_\theta(\mathbf{x}|\mathbf{z}) = f(\mathbf{x}; g_\theta(\mathbf{z})) \tag{4.15}$$

where g_θ is a NN parameterised by θ, and f is typically a Gaussian or Bernoulli distribution depending on the nature of the data.

The Reparameterisation Trick: A key challenge in training VAEs is that sampling from $q_\phi(\mathbf{z}|\mathbf{x})$ is non-differentiable, preventing straightforward backpropagation. Kingma and Welling [102] addressed this with the reparameterisation trick, which reformulates the sampling process as:

$$\mathbf{z} = \boldsymbol{\mu}_\phi(\mathbf{x}) + \boldsymbol{\sigma}_\phi(\mathbf{x}) \odot \boldsymbol{\epsilon} \tag{4.16}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is a standard multivariate normal distribution with the same dimensions as the latent space. This maintains the stochastic nature while simultaneously allowing backpropagation.

4.7.2. KL Divergence and the VAE Loss Function

The VAE is trained to maximise the Evidence Lower BOund (ELBO) which is a lower bound on the log-likelihood of the data:

$$\mathcal{L}_{\text{ELBO}}(\mathbf{x}) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] - D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z})) \tag{4.17}$$

where D_{KL} is the Kullback-Leibler divergence, and $p(\mathbf{z})$ is a prior distribution over the latent space, typically chosen as $\mathcal{N}(\mathbf{0}, \mathbf{I})$.

The KL divergence term measures the dissimilarity between the encoder's distribution $q_\phi(\mathbf{z}|\mathbf{x})$ and the prior $p(\mathbf{z})$:

$$D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z})) = \int q_\phi(\mathbf{z}|\mathbf{x}) \log \frac{q_\phi(\mathbf{z}|\mathbf{x})}{p(\mathbf{z})} d\mathbf{z} \tag{4.18}$$

For the case of Gaussian distributions, this has a closed-form solution:

$$D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z})) = \frac{1}{2} \sum_{j=1}^{d} \left(\boldsymbol{\mu}_\phi(\mathbf{x})_j^2 + \boldsymbol{\sigma}_\phi^2(\mathbf{x})_j - \log \boldsymbol{\sigma}_\phi^2(\mathbf{x})_j - 1 \right) \tag{4.19}$$

The KL divergence term is primarily a regulariser, encouraging the encoder to produce latent distributions close to the standard normal prior. This has several important implications:

1. It ensures a well-structured, continuous latent space where similar inputs inhabit similar regions of the latent space.
2. It prevents the model from simply memorising the training data, leading to meaningful representations that generalise well.
3. It facilitates smooth interpolation between different data points, laying the ground for generative modelling by sampling from the latent space.

The complete VAE loss function can thus be written as:

$$\begin{aligned} \mathcal{L}_{\text{VAE}}(\mathbf{x}) &= \mathcal{L}_{\text{recon}}(\mathbf{x}) + \mathcal{L}_{\text{KL}}(\mathbf{x}) \\ &= -\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] + D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z})) \end{aligned} \tag{4.20}$$

where $\mathcal{L}_{\text{recon}}$ represents the reconstruction loss and $\mathcal{L}_{\text{KL}}$ is the KL divergence term. In practice, these components are often balanced with a hyperparameter β [105]:

$$\mathcal{L}_{\text{VAE}}(\mathbf{x}) = \mathcal{L}_{\text{recon}}(\mathbf{x}) + \beta \cdot \mathcal{L}_{\text{KL}}(\mathbf{x}) \tag{4.21}$$

This formulation of the so-called β-VAE, which is *also used* in this thesis, allows the user to control the trade-off between reconstruction fidelity and latent space regularity, which can be particularly valuable in those applications where this balance can have a significant impact on performance [106].

4.7.3. Applications in Particle Physics and Anomaly Detection

VAEs combine a reconstruction objective with a Kullback-Leibler term that drives the approximate posterior $q_\phi(\mathbf{z}\,|\,\mathbf{x})$ towards a simple Gaussian prior $p(\mathbf{z}) \simeq \mathcal{N}(0,1)$ [102]. The resulting *continuous* latent space, comprising as many dimensions as the user chooses, can be described as a manifold of Gaussian nature. Fundamentally, it provides several advantages over conventional autoencoders for model-agnostic searches at the LHC:

- Robust anomaly scoring: Because the latent space is explicitly regularised, the corresponding anomaly scores are less sensitive to statistical fluctuations. This has been shown to improve outlier rejection in previous studies [106, 107].
- Generative modelling and uncertainty quantification: The decoder maps latent points back to the same set of input features, which could for example be detector-level observables. In principle, this would allow fast, physics-aware data augmentation and the construction of likelihood ratios directly from generated samples. These features facilitate side-band extrapolations and background estimates where data is limited [108]. However, improved algorithms such as Generative Adversarial Networks (GANs) [109] have largely superseded VAEs in performance when it comes to purely generative tasks.
- Interpretable latent representations: The smooth latent manifold with its Gaussian nature enables post-training inspection of the regions most correlated with larger anomaly scores.

VAEs have been successfully used in several HEP analyses to gain further insights into the challenging task of anomaly detection, a few of which are listed below.

- The pioneering study of Ref. [106] established the one-sided VAE threshold test to flag anomalous events in CMS open data.
- An unsupervised VAE successfully identified Soft Unclustered Energy Patterns (SUEP) in $H \to$ hidden scenarios, probing percent-level branching ratios at the HL-LHC [110].
- Mass-decorrelated VAEs achieved competitive performance for anti-QCD tagging while simultaneously reducing sensitivity to outliers [107].
- Resource-optimised VAEs designed to run on FPGAs have been integrated in the CMS L1 global trigger for real-time anomaly detection [111] at an event rate of 40MHz.
- Improvements on top of VAEs such as normalising-flow layers used alongside VAEs improve separation without sacrificing the closed-form latent prior [108], while variational *quantum* circuits have been explored as encoders on current NISQ hardware [8].

These developments underscore the versatility of VAEs and their central role in forthcoming model-independent searches at the LHC and beyond.

5. Model Agnostic Search for Dijet Resonances with VAEs at $\sqrt{s} = 13$ TeV

5.1. Introduction to the CMS-EXO-22-026 Analysis

This section introduces the **CMS EXO-22-026** [51, 112] Analysis, performed as part of the internally-named CMS Anomaly Search Effort (CASE). As a member of the analysis team, I made major contributions to the development and implementation of the Variational Autoencoder-Quantile Regression (VAE-QR) method, which was one of five methods introduced as part of this novel analysis, the others being Classification Wit**h**out L**a**bels (CWoLa) Hunting, Tag 'N Train (TNT), Classifying Anomalies THrough Outer Density Estimation (CATHODE) and QUasi Anomalous Knowledge (QUAK).

The following individuals were involved in the EXO-22-026 analysis:

- **CERN:** Maurizio Pierini, Kinga Anna Wozniak [113]
- **ETH Zurich:** Thea Klaeboe Aarrestad
- **Fermi National Accelerator Laboratory (Fermilab):** Oz Amram, Jennifer Ngadiuba, Irene Zoi
- **University of Hamburg:** Gregor Kasieczka, Louis Jean Moureaux, Tobias Quadfasel, Tore von Schwartz, Manuel Sommerhalder [114], Chitrakshee Yede
- **Imperial College London:** Benedikt Maier
- **Johns Hopkins University:** Petar Maksimovic
- **Karlsruhe Institute of Technology (KIT):** Aritra Bal (*author of this thesis*)
- **Massachusetts Institute of Technology (MIT):** Philip Harris, Samuel Bright-Thonney [115], William Patrick McCormack, Sang Eon Park
- **Rutgers University:** David Shih

Searches for new physics in hadronic final states constitute an essential component of the experimental program at the LHC. While traditional resonance searches have proven effective in probing various extensions of the SM, they are often optimised for specific decay channels, limiting their sensitivity to a broader range of potential signals. Therefore, the **CMS-EXO-22-026** analysis is the first model-agnostic search within the CMS Collaboration that looks for narrow resonances in the dijet final state within the mass range of 1.8 – 6 TeV, focusing on topologies where a heavy resonance A decays into two intermediate particles B and C, which subsequently decay hadronically. Due to their significant Lorentz boost, the decay products of these intermediate particles are captured within large-radius jets exhibiting distinctive substructure patterns which the algorithms presented therein are expected to be sensitive

to. Unlike conventional searches, this analysis employs multiple complementary ML-based paradigms for anomaly detection, these being unsupervised (VAE-QR), weakly supervised (CWoLa,TNT,CATHODE), and semi-supervised (QUAK), to identify jets with non-QCD-like substructure, thereby enhancing sensitivity. This analysis looks at a wide variety of signal topologies ranging from simple 1+2 prong configurations to complex 6+6 prong signatures, offering a powerful and versatile framework for discovering new physics that might escape targeted searches focused on a single signal model.

Table 5.1.: Table of signal resonances of the form $A \to BC$, showing the number of prongs in the jets arising from the decays of the B and C daughter particles (with the bold masses of the daughter particles showing the signal models used for benchmarking in Section 5.9.)

$C \downarrow$ \ $B \to$	2 prongs	3 prongs	5 prongs	6 prongs
1 prong	$\mathrm{Q}^* \to \mathrm{qW}$ $m_{Q^*} = [2, 3, 5]$ TeV $m_W = [25, 80, 170, 400]$ GeV			
2 prongs	$\mathrm{X} \to \mathrm{YY}'$ $m_{\mathrm{X}} = [2, 3, 5]$ TeV $m_{\mathrm{Y}} = [25, 80, 170, \mathbf{400}]$ GeV $m_{\mathrm{Y}'} = [25, 80, 170, \mathbf{400}]$ GeV			
3 prongs		$\mathrm{W}' \to t\mathrm{B}'$ $m_{W'} = [2, 3, 5]$ TeV $m_{B'} = [25, 80, 170, \mathbf{400}]$ GeV		
4 prongs	$\mathrm{W_{KK}} \to \mathrm{WR} \to \mathrm{WWW}$ $m_{W_{\mathrm{KK}}} = [2, 3, 5]$ TeV $m_R = [\mathbf{170}, 400]$ GeV			
5 prongs			$\mathrm{Z}' \to \mathrm{T}'\mathrm{T}' \to t\,\mathrm{Z}\,t\,\mathrm{Z}$ $m_{Z'} = [2, 3, 5]$ TeV $m_{\mathrm{T}'} = [400]$ GeV	
6 prongs				$\mathrm{G}_{kk} \to \mathrm{HH} \to tttt$ $m_{\mathrm{Y}} = [2, 3, 5]$ TeV $m_{\mathrm{H}} = [\mathbf{400}]$ GeV

The VAE-QR method represents the only fully unsupervised anomaly detection approach employed in this analysis, as it does not require any labelled jets for training, nor does it need to make use of MC signal simulation at any point of its workflow. This technique is distinguished by its use of low-level features, these being the kinematic vectors of the 100 highest p_{T} jet constituents. The method employs a VAE trained exclusively on data from a control region dominated by QCD jets, which learns to compress and reconstruct these jets effectively. When presented with anomalous jets not represented in its training set, the reconstruction performance of the VAE degrades, producing higher reconstruction loss values that can serve as effective anomaly scores. To prevent sculpting of the dijet invariant mass

(which is elaborated upon in the next paragraph) spectrum that would complicate background estimation and reduce signal efficiency, a DNN based QR is used to decorrelate the VAE anomaly score from the dijet invariant mass m_{jj}. This decorrelation preserves the shape of the m_{jj} distribution while maintaining sensitivity to potential signals across all jet substructure types and across a wide range of m_{jj}, giving VAE-QR the broadest sensitivity among all methods presented in this analysis.

The dijet invariant mass, denoted as m_{jj}, is a fundamental kinematic quantity in searches for new physics phenomena in dijet systems. This quantity is calculated from the four-momentum vectors of the two leading jets in an event and provides a Lorentz-invariant measure of the total energy-momentum of the system. The m_{jj} spectrum is crucial for identifying resonances that may indicate the presence of new, potentially BSM, particles decaying to jet pairs. At the CMS experiment, dijet mass distributions are extensively used in searches for physics beyond the SM, where deviations from the smoothly falling QCD multijet background could signal the presence of new physics processes. Given the four-vectors of the two jets $(E_1, \vec{p}_1)$ and $(E_2, \vec{p}_2)$, the dijet invariant mass can be written as:

$$m_{jj} = \sqrt{(E_1 + E_2)^2 - (\vec{p}_1 + \vec{p}_2)^2} \tag{5.1}$$

which in the limit of low jet mass $m_{jet} << p_{T,jet} \cosh(\eta_{jet})$ reduces to:

$$m_{jj} = \sqrt{2 p_T^{(1)} p_T^{(2)} (\cosh(\eta_1 - \eta_2) - \cos(\phi_1 - \phi_2))} \tag{5.2}$$

Following the selection of events with anomalous jet substructure, a search for resonant signatures is performed over the resulting m_{jj} spectrum. With no significant excesses observed, the analysis establishes exclusion limits on a variety of signal models (see Table 5.1) spanning different jet substructure complexities (from 2+2 prongs to 6+6 prongs). Notably, for many of these benchmark signals, this analysis presents the first exclusion limits ever set at the LHC, demonstrating the VAE-QR method's effectiveness as a fully model-agnostic approach for new physics searches.

5.2. Description of the datasets used

The analysis presented in this thesis utilises pp collision data recorded at the CMS detector at the LHC during the period 2016-18, known commonly as the Run 2 era. The dataset contains events recorded at a center-of-mass energy $\sqrt{s} = 13$ TeV, corresponding to a total integrated luminosity of 138 fb^{-1}.

5.2.1. Data Quality and Selection

To ensure high-quality data appropriate for physics analysis, strict criteria were applied to select only periods when all CMS subdetectors were functioning optimally. These criteria are detailed in the so-called *golden JSON* files, which enumerate the validated luminosity sections suitable for offline analysis [116] within the CMS Collaboration. The golden JSON files for each data-taking year, along with their respective integrated luminosities, are detailed in Table 5.2.

Table 5.2.: Golden JSON files showing the integrated luminosity by year.

Year	Golden JSON file name	$\mathcal{L}_{\text{int}}$ [fb^{-1}]
2016	Cert_271036-284044_13TeV_Legacy2016_Collisions16_JSON.txt	36.4
2017	Cert_294927-306462_13TeV_UL2017_Collisions17_GoldenJSON.txt	41.5
2018	Cert_314472-325175_13TeV_Legacy2018_Collisions18_JSON.txt	59.8
Total		137.6

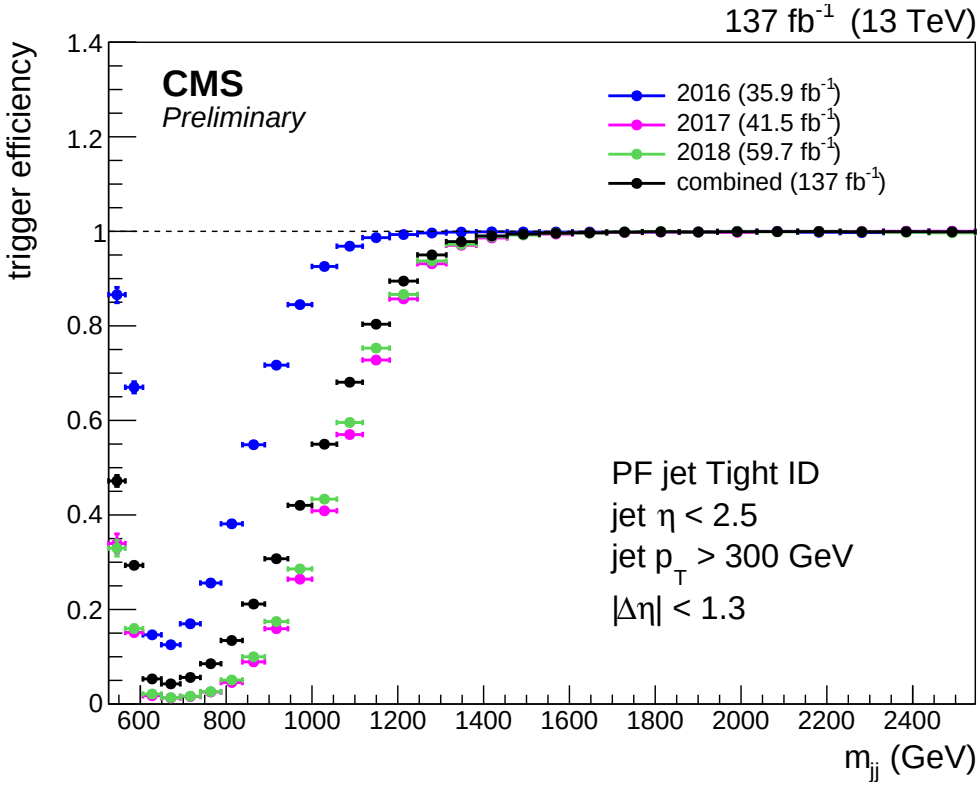


Figure 5.1.: Trigger efficiency vs dijet invariant mass m_{jj} for each year, and also for the entire Run 2. The increase in trigger efficiency in the low m_{jj} region, particularly for $m_{\text{jj}} < 600$ GeV is due to the high $p_{\text{T}} > 300$ GeV requirement on the leading and subleading jets in the event. As can be seen in Equation 5.2, this is possible only when the jets are aligned in roughly the same direction, with transverse momentum conservation requiring the presence of additional jets that are aligned oppositely. H_{T} being the scalar sum of jet transverse momenta in the event, events in this region of phase space end up activating the trigger (figure from CMS internal Analysis Note AN-22-026).

5.2.2. JetHT Dataset and Trigger Selection

This analysis specifically uses the **JetHT** primary dataset, which contains events passing trigger selections designed to select highly energetic jets. The JetHT dataset is particularly suitable for analyses targeting rare jet-based signatures. Events in this dataset have satisfied at least one of a number of triggers designed to efficiently collect collision events containing at least one high-p_{T} jet, or possessing a high scalar sum of jet transverse momenta (H_T).

This is achieved by using a combination of jet p_{T} and event H_T triggers employed during the three years of Run 2. Table 5.3 summarises these triggers, which vary slightly across different data-taking periods to accommodate changing detector and beam conditions. The triggers include both PF-based high-level trigger paths targeting high p_{T} jets, and total event H_{T} triggers which select events that have a large scalar sum of jet transverse momenta.

To ensure full trigger efficiency in the analysis phase space, a dedicated efficiency study was conducted using the Single Muon dataset with muon trigger paths (HLT_IsoMu27 and HLT_Mu50) that are statistically independent from the hadronic triggers used in this analysis. Based on this study, it was observed that the combined trigger efficiency reaches a plateau of $> 99\%$ for m_{jj} above 1455 GeV in 2017 and 2018 data, while for 2016 data the plateau begins at approximately 1181 GeV (see Figure 5.1 and also Figure B.1 in Appendix B.1). To maintain consistent selection criteria across all years, an analysis threshold of $m_{jj} > 1455$ GeV is applied, ensuring that only events from the region of full trigger efficiency are utilised.

Table 5.3.: Jet-based triggers used to collect data online during the three data-taking years.

Period	H_T trigger	jet p_T trigger
2016 runs B-G	HLT_PFHT800	HLT_PFJet450
2016 run H	HLT_PFHT900	HLT_PFJet450 HLT_AK8PFJet450
2017 runs B-F	HLT_PFHT1050	HLT_AK8PFJet500
2018 runs A-D	HLT_PFHT1050	HLT_AK8PFJet500

5.2.3. Data Processing

Starting from the CMS Ultra Legacy (UL) campaign, this analysis requires the production of custom NanoAOD samples that extend the standard content to include the kinematics of individual jet PFCands, which are essential inputs to the VAE-QR. These custom samples were generated using the JetMET PFNano producer. For the 2016 data, MiniAODv1 served as input, while MiniAODv2 was used for the 2017 and 2018 datasets. All samples were processed using the `CMSSW_10_6_20` release with the `106X_dataRun2_v32` global tag, ensuring consistent reconstruction algorithms and calibrations across the entire dataset.

5.2.4. Noise Filtering and Selection of Vertices

To maintain the integrity of the analysis, it is crucial to eliminate events affected by misreconstructed MET resulting from various instrumental effects. These could include anomalies detected in the HCAL or irregular energy deposits in the ECAL. Following the recommendations established by the MET Physics Object Group (POG) of the CMS Collaboration, this analysis uses the event-level noise filters listed below:

- `goodVertices`
- `globalSuperTightHalo2016Filter`
- `HBHENoiseFilter`
- `HBHENoiseIsoFilter`
- `EcalDeadCellTriggerPrimitiveFilter`
- `BadPFMuonFilter`
- `BadPFMuonDzFilter`

- `eeBadScFilter`
- `CSCTightHaloFilter` (2016 only)
- `ecalBadCalibFilter` (2016 excluded)

In addition to these noise filters, the analysis strategy enforces stringent vertex quality requirements to ensure proper event reconstruction. Each event must contain at least one primary vertex reconstructed within a 24 cm window along the beam axis, with a transverse distance from the nominal pp interaction region of less than 2 cm. When multiple vertices satisfy these criteria, the vertex with the highest total p_{T}^2 (summed over all associated tracks) is designated as as the primary one.

It is worth noting that during the 2018 data-taking period, several HCAL modules experienced functional failures, causing erroneous jet energy measurements in the affected regions. In the context of this analysis, it is essential to minimise contamination from detector malfunction, and so such events from this period where one of the two hardest jets falls within the affected regions $\phi \in (-1.57, -0.87)$ and $\eta \in (-2.5, -1.3)$ are explicitly rejected, using a so-called *veto*. This cautious approach results in an approximately 1% decrease in signal efficiency for 2018 data.

5.2.5. Event Preselection

The event reconstruction is carried out using the PF algorithm described in Section 3.9. Jets are reconstructed from these PF candidates using the anti-k_{T} clustering algorithm [39] with a distance parameter of $R = 0.8$ (AK8 jets), as implemented in the FastJet package [42], though for the specific case of VAE-QR, an additional reclustering step becomes necessary as described in Section 5.3.1.

To mitigate pileup effects, which could impact jet substructure analysis, the Pileup Per Particle Identification (PUPPI) algorithm [65] is used. In short, this approach incorporates local shape information of charged pileup particles, event-level characteristics, along with information from the tracker to compute a scale factor that rescales the p_{T} of each neutral and charged PF candidate. The algorithm computes a probability for each particle, as to whether or not it originates from a pileup interaction. Charged particles not associated with the primary vertex receive a weight of zero by default.

All AK8 jets must satisfy tight jet identification criteria as established by the JetMET POG for Run 2 data. These criteria are summarised in Table 5.4 and include requirements on neutral hadron fraction, neutral electromagnetic fraction, number of constituents, charged hadron fraction, and charged multiplicity. Jets are corrected for nonlinearities in p_{T} and rapidity using standard jet energy corrections developed by the CMS Collaboration.

The signal region is constructed by selecting events containing at least two jets with $p_{\mathrm{T}} > 300\,\mathrm{GeV}$, each passing the Tight Jet ID requirements and satisfying $|\eta| < 2.5$. The two highest-p_{T} jets meeting these criteria are designated as the candidate *dijet* system. To reduce QCD multijet background, we further require the two jets to have a separation of $|\Delta\eta_{jj}| < 1.3$ and a dijet invariant mass $m_{\mathrm{jj}} > 1455\,\mathrm{GeV}$, for the reason already mentioned in Section 5.2.2. The complete set of preselection criteria is summarised in Table 5.5:

Table 5.4.: Jet ID criteria for AK8 PUPPI jets with $|\eta| < 2.5$.

Quantity	Threshold
Neutral Hadron Fraction	< 0.90
Neutral EM Fraction	< 0.90
Number of Constituents	> 1
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0

Table 5.5.: Summary of event preselection criteria defining the signal region for the dijet anomaly search.

Preselection Criterion	**Requirement**
Jet algorithm	anti-k_T, $R = 0.8$ (AK8)
Jet quality	PF Tight ID applied
Jet η	$\|\eta\| < 2.5$
Leading jet p_T	$p_T > 300\,\text{GeV}$
Pseudorapidity separation	$\|\Delta\eta_{jj}\| < 1.3$
Dijet invariant mass	$m_{jj} > 1455\,\text{GeV}$

It is important to note that while these criteria define the signal region in which the search is performed, the VAE of VAE-QR is actually trained on jets sampled from a control region defined in Section 5.4. This ensures that the model learns an accurate representation of the QCD multijet background without contamination from potential signal, while not being biased by the kinematics of the aforementioned control region.

5.2.6. Simulation of MC Samples

While this analysis employs a data-driven approach for background estimation, thus not strictly requiring background MC simulation, background MC samples were nevertheless produced to refine and validate the analysis strategy. The dominant background process is the QCD multijet, simulated at leading order (LO) with Pythia 8.2 [117, 118, 119], while secondary SM backgrounds include:

- V+jets (where V = W, Z) simulated at LO with MadGraph5_aMC@NLO v2.6.5 [120],
- $t\bar{t}$ pair production and single top quark production generated at next-to-leading order (NLO) with Powheg v2 [121, 122].

All processes employed the NNPDF 3.1 [123] parton distribution functions. The parton shower was simulated using Pythia 8.2, with the CP5 tune [124]. Detector interactions were simulated with Geant4 [125]. The samples were produced for each of the four data-taking periods (2016 preVFP, 2016 postVFP, 2017, 2018) with appropriate phase space binning. To mitigate issues with event weighting typical in conventional analyses, a realistic number of events was sampled according to effective target luminosity, process cross-section, and selection efficiency as follows:

$$N_{\text{sampled}} = \mathcal{L}_{\text{eff}} \times \sigma \times \varepsilon_{\text{preselection}} \tag{5.3}$$

This resulted in a total integrated luminosity of 26.8 fb^{-1} for the background dataset, which is approximately five times smaller than what was recorded by the CMS Detector during the Run 2 data-taking period.

Signal MC samples, on the other hand, are crucial for setting exclusion limits on the theoretical models under investigation, many of them for the first time as already mentioned. As summarised in Table 5.1, this analysis targets a general resonance A decaying to two other generic resonances B and C in a single search framework. The focus is on the scenario where these secondary resonances decay hadronically, resulting in final states with two jets. Various possibilities regarding the masses of the resonances and the number of hadrons produced in the subsequent decays of B and C are considered. Specifically, the analysis explores scenarios with 2 to 6 hadrons in the decay chains, with resonance A having mass values of [2, 3, 5] TeV . For cases where B and C are massive particles, masses of [25, 80, 170, 400] GeV are considered. In scenarios involving a known SM particle (such as the W, Z, H bosons, or the top quark), only the mass of the unknown particle is scanned. Overall, several signal simulations are necessary.

The simulated events for these signal models are generated at LO in QCD using MadGraph5_aMC@NLO 2.6.5 with the MLM matching scheme. The hadronisation and parton showering processes are simulated using Pythia `8.226` for 2016 samples and `8.230` for 2017 and 2018 samples. For the 2016 sample production, the `NNPDF3.0_lo_as_0130 LO` PDF set is employed, while the `NNPDF3.1 NNLO` set is used for 2017 and 2018 production. All resonances in this study are generated with narrow width, though it should be noted that the VAE-QR approach is *not inherently limited* by this constraint and is shown to potentially exhibit sensitivity to broad resonances as well. All MC sample production was carried out within the official CMS MC production framework of the UL campaign.

5.2.6.1. Choice of Signal Models

The simulation of signal samples was described in the previous section. This section provides an overview of each signal model considered in this analysis. The complete set of signal models evaluated in this study is presented in Table 5.1.

It is emphasised here that the VAE-QR approach is fundamentally model-agnostic and requires minimal assumptions about the signal characteristics. The signal models presented here represent only a small subset of the potential models this methodology could be applied.

Six distinct signal models with varying topologies were considered:

- $Q^* \rightarrow qW' \rightarrow 3q$: The first signal model exhibits a 1+2 prong topology, consisting of an excited quark resonance (Q^*) decaying into a quark and W′ boson [126, 127]. The W′ boson subsequently decays into two light quarks with flavour composition identical to the SM W boson.
- $X \rightarrow YY' \rightarrow 4q$: The second signal model features a 2+2 prong topology and comprises a heavy resonance X decaying to two resonances, Y and Y′. Each of these secondary resonances then decays into two light quarks.
- $W' \rightarrow B't \rightarrow bZt$: The third signal model presents a 3+3 prong topology and involves a W′ boson decaying into a vector-like quark (B′) and a top quark [128]. The B′ subsequently decays into a bottom quark and a Z boson.
- $W_{KK} \rightarrow RW \rightarrow 3W$: The fourth signal model demonstrates a 2+4 topology and consists of a W'_{KK} boson decaying into a radion (R) and a W boson [129], with the radion subsequently decaying to two W bosons.

- $Z' \rightarrow T'T' \rightarrow tZtZ$: The fifth signal model displays a 5+5 topology and comprises a Z' boson decaying into two vector-like quarks, T' [130, 131]. These vector-like quarks then decay into a top quark and a Z boson.
- $G_{KK} \rightarrow HH \rightarrow 4t$: The sixth signal model exhibits a 6+6 topology and involves a heavy spin-2 Randall-Sundrum graviton (G_{KK}) decaying to two lighter Higgs-like scalars (H) [132]. Each lighter Higgs-like scalar subsequently decays into a top quark-antiquark pair.

The $W' \rightarrow B't \rightarrow bZt$ process was previously investigated by the CMS Collaboration [133]. However, the present analysis explores a different parameter space where the B' is sufficiently light compared to the W', with its decay products being merged into a single jet.

The $W_{KK} \rightarrow RW \rightarrow 3W$ process has also been searched for in the past by the CMS Collaboration [134, 135]. Therefore, the inclusion of this signal facilitates a direct comparison between the sensitivity of the unsupervised and model-agnostic method presented here, to that of an optimised dedicated search.

The content of the preceding sections should be sufficient for the reader to gain an understanding of the model-agnostic nature and the broad goal of this analysis, along with the datasets used to carry it out. In the following sections, the technical details and analysis strategy of the VAE-QR method, and thereafter the results obtained, will be described in detail.

5.3. Setting up the VAE for Anomaly Detection at the CMS Experiment

The mathematical formulation of a VAE has been thoroughly discussed in Section 4.7, providing the theoretical framework for this analysis. In this section, the specific technical details are presented. To recap, VAEs serve as powerful tools for anomaly detection by learning to encode and reconstruct the statistical features of *normal* data. When faced with the task of reconstructing data that possibly originates outside the learned distribution, its reconstruction performance suffers, allowing this *reconstruction* loss to serve as an anomaly metric. The key advantage of this approach lies in its model-agnostic nature, since the VAE requires no explicit assumptions about potential anomalous signatures, allowing for sensitivity to a wide range of BSM phenomena without being optimised for any specific signal hypothesis. The implementation details presented in this section reflect the architectural choices, hyperparameter optimisation, and practical considerations that adapt the general VAE framework to the specific challenge of detecting anomalies in dijet resonances at the LHC.

5.3.1. Input Data and Features

The VAE-QR method is distinct from the four other approaches presented in this analysis, with regards to the input representation it relies upon. While the other methods utilise high-level features constructed to already have a high discriminative power, the VAE-QR operates directly on minimally processed, low-level jet constituent PFCand data. Specifically, the inputs to the VAE-QR algorithm are the momentum components (p_x, p_y, p_z) of the 100 highest p_T jet constituents. This approach deliberately preserves the fundamental physical information without introducing higher-level abstractions that might inadvertently lead to information

loss due to compression. The nature of the input datasets is already elaborated upon in Section 5.2.

Initial experimentation included several training iterations utilising alternative feature representations, particularly the $(p_T, \eta_{rel}, \phi_{rel})$ for each jet constituent, where η_{rel} and ϕ_{rel} denote the pseudorapidity and azimuthal angle relative to the jet axis, respectively. However, these configurations consistently yielded suboptimal results. This underperformance can be attributed to the presence of sharp distributional boundaries in the η-ϕ space, which introduce discontinuities that adversely affect the learning process of the VAE. The momentum components, conversely, exhibit smoother distributions that facilitate more effective latent space encoding.

As already mentioned in Section 5.2, this analysis looks for resonant anomalies in dijet events, which means that each event contains at minimum two jets. Throughout this discussion, the leading and subleading jets will be designated as J_1 and J_2, respectively, based on their p_T ordering. A critical aspect of the methodology is that training proceeds on a **per-jet** basis rather than per-event. Each event is split into its constituent jets, which are then processed independently by the VAE. Consequently, the autoencoder operates without an awareness of event-level correlations or the possibility that two jets in the training set might originate from the same physical event. This jet-centric approach aligns with the physics goal of identifying anomalous jet substructure patterns regardless of an event-level association. The per-jet anomaly scores so obtained are later combined to obtain an event-level anomaly score following the methodology outlined in Equation 5.11.

The input data is structured as a three-dimensional tensor with dimensions $(N_{batch}, 100, 3)$, where N_{batch} represents the batch size during training, 100 corresponds to the maximum number of constituents considered per jet, and 3 represents the momentum three-vector (p_x, p_y, p_z) for each constituent. The following transformations are then necessary to convert from (p_T, η, ϕ) to (p_x, p_y, p_z):

$$p_x = p_T \cos(\phi) \tag{5.4}$$

$$p_y = p_T \sin(\phi) \tag{5.5}$$

$$p_z = p_T \sinh(\eta) \tag{5.6}$$

For jets containing fewer than 100 constituents, the tensor is zero-padded to maintain consistent dimensionality across all samples. While it is shown in Section 5.3.3 that the choice of reconstruction loss function is permutation invariant, it was observed during training that *reclustering* the anti-k_T ($R = 0.8$) jets using the C/A algorithm and using the subsequent PFCand ordering, leads to moderate improvements in performance, which was then taken as the standard. All in all, this standardised representation ensures uniform processing through the neural network architecture while preserving the essential kinematic information of each jet.

5.3.2. Architecture

The backbone of the VAE used in VAE-QR is the Convolutional Neural Network (CNN) [136, 137]. The architecture in this analysis implements a VAE with a 12-dimensional latent space, designed to process PFCand data. As mentioned earlier, the input is an array of particle

momenta vectors. The two major components of the VAE, its encoder and decoder, are summarised below.

The encoder network performs a series of transformations:

- Normalisation: This layer uses the mean μ and standard deviation σ of the dataset to perform the scaling:
$$x_i \rightarrow \frac{x_i - \mu}{\sigma} \tag{5.7}$$
so that the input batches have a mean μ of 0 and a standard deviation σ of 1.
- **Convolutions**:
  - LambdaExpand: Before the convolutional kernels can be used on the input, it is necessary to add an extra dimension to the input, expanding the shape of each event array to $(100, 3, 1)$, this being necessary to indicate the number of channels in the input, which in this case is just 1.
  - 2DConv: The first convolution operation utilises a $1 \times 3 \times 12$ kernel, these being the kernel height, width and output channels respectively. The kernel height of 1 ensures that it *slides* along the jet constituents, applying a $1D$ convolution, and applies the full width of the kernel (3) to match the remaining dimensions of the input tensor ((3×1)). There are effectively 12 different kernels, spanning the entire width of the input at each position, processing the three momentum components simultaneously, while preserving the sequential structure along the first dimension. This transforms the single input channel into 12 distinct feature maps, resulting in a final output tensor of shape $(N \times 100 \times 1 \times 12)$. Each of the twelve output channels represents a different learned outcome across the three-dimensional momentum space.
  - LambdaSqueeze: Before moving to the 1D convolutional operation, this operation is necessary to get rid of the middle dimension which is no longer required, resulting in a tensor of shape $(100, 12)$.
  - 1DConv: The first one-dimensional convolution applies 16 filters with a kernel size of 3 to the squeezed tensor of shape $(100, 12)$. Each filter spans 3 consecutive jet constituents across all 12 feature channels from the previous layer. This operation effectively creates a sliding window of size 3 that moves across the sequence of 100 constituents, generating 16 new feature maps. The resulting output tensor has dimensions $(100 - 3 + 1, 16) = (98, 16)$, where the reduction in the first dimension occurs due to the kernel width.
  - 1DConv: The second one-dimensional convolution further processes the feature maps by applying 20 filters with a kernel size of 3 to the tensor of shape $(98, 16)$. Each filter spans 3 consecutive positions across all 16 feature channels. This operation enhances the model's ability to capture more complex patterns and hierarchical representations within the jet data. The output tensor from this layer has dimensions $(98 - 3 + 1, 20) = (96, 20)$, again reflecting the reduction in sequence length due to the convolution operation with valid padding. This sequential application of convolutional layers enables the network to build increasingly abstract representations of the jet substructure.
- **Final Data Compression:**

  - Pooling: A 1D average pooling layer is used to collate information in preparation for flattening
  - Flatten+Dense: A single flattening layer is followed by multiple dense fully connected layers to reduce dimensionality to a tensor of shape 48, following which two separate dense layers are used to create the latent space parameterisation of the VAE in terms of the mean μ_{latent} and standard deviation σ_{latent}.

- **Latent Space Parameterisation:**
  - Outputs two vectors μ_{latent} and σ_{latent}
  - Generates a **12-dimensional** latent representation

The decoder network is effectively the inverse of the encoder, as it performs the reconstruction process to get back an output tensor that matches the input dimensionality $(N, 100, 3)$ using the following transformations:

- **Latent Space Sampling:** This uses the important reparameterisation trick of Equation 4.16 to sample from the latent distribution and obtain a tensor of shape 12.
- Dense: Multiple dense fully connected layers progressively increase the dimensionality of the samples drawn from the latent space, laying the ground for the upsampling and transpose convolutional operations that restore the original dimensionality.
- **Upsampling and Transpose Convolutional Layers:**
  - Upsampling1D: reconstructs the original input dimensionality
  - 2x 1DConv + 2DConv transpose: These 3 operations, followed by a squeezing operation bring the tensor to a shape of $(N, 100, 3)$.
- UnNormalisation: The final missing ingredient in the VAE recipe is the reversal of the standard normalisation using the inverse of the transform of Equation 5.7:

$$x_i \rightarrow \mu + \sigma \cdot x_i \tag{5.8}$$

The VAE therefore learns a compressed 12-dimensional representation of the jet PFCand momentum space, which it then uses to reconstruct its input. The architecture is shown as a schematic in Figure 5.2.

5.3.3. Loss Function

The reconstruction loss employed in this work is the 3D Chamfer Distance [138], used extensively for point cloud reconstruction tasks. This metric is particularly suitable for the current application, as the input data consists of a point cloud of PFCands within a jet, where both the number and ordering of points can vary between samples.

The Chamfer Distance computes a symmetric measure between two point clouds by finding, for each point in one set, the nearest point in the other set, and vice versa. Mathematically, this loss $\mathcal{L}_{\text{RECO}}$ is defined as:

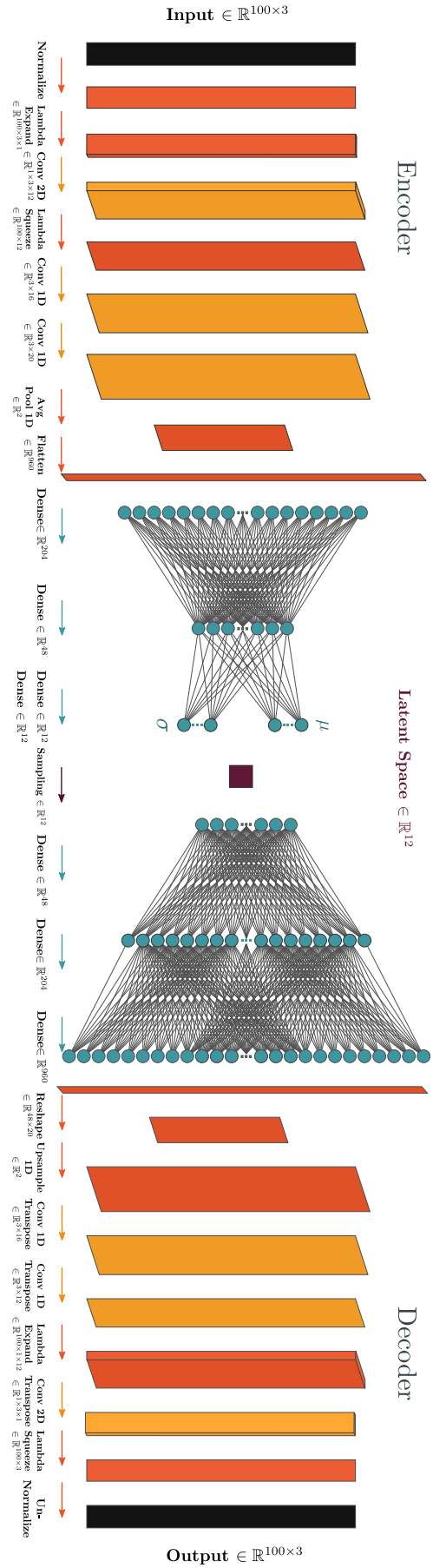


Figure 5.2.: Architecture of the VAE

$$\mathcal{L}_{\text{RECO}} = \sum_{i \in \text{input}} \min_{j} \left(x_i - \hat{x}_j\right)^2 + \sum_{j \in \text{output}} \min_{i} \left(x_i - \hat{x}_j\right)^2 . \tag{5.9}$$

A key advantage of this loss function is its permutation invariance, as it calculates all pairwise distances between input and reconstructed jet PFCands rather than requiring a fixed ordering or one-to-one correspondence between points. This property is essential when working with jets, where the number and arrangement of constituents can vary significantly from event to event.

Furthermore, following the β-VAE formulation introduced by [105] and given in Equation 4.21, a weighted Kullback-Leibler divergence term is added to the loss function. This additional component ensures that the latent space approximates a standard Gaussian distribution, thus striking a balance between maintaining a Gaussian latent manifold and achieving effective jet reconstruction. The hyperparameter β, which is set equal to 0.5 controls the trade-off between these two objectives.

The final loss function per-jet is:

$$\mathcal{L}_{\text{RECO}} = \sum_{i \in \text{input}} \min_{j} \left(x_i - \hat{x}_j\right)^2 + \sum_{j \in \text{output}} \min_{i} \left(x_i - \hat{x}_j\right)^2 + \frac{1}{2} \cdot D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z})) \tag{5.10}$$

where $q_\phi(\mathbf{z}|\mathbf{x})$ and $p(\mathbf{z})$ are the latent space distributions and Gaussian priors respectively.

It is to be noted that the VAE-QR is an event-level anomaly detection algorithm. However, since the training is on a per-jet basis, it becomes necessary to combine the per-jet loss of Equation 5.10, as follows:

$$\mathcal{L}_{\text{EVENT}} = \min\left(\mathcal{L}_{\text{J}_1}, \mathcal{L}_{\text{J}_2}\right) \tag{5.11}$$

5.3.4. Technical Aspects of the Network Training

The VAE is implemented and trained using the TensorFlow [139] package of the Python programming language [140]. The network is trained on a dataset consisting of 1 million collision events, which translates to approximately 2 million jets, since each event contributes two jets for analysis. For validation purposes, a separate dataset of 250,000 events (equivalent to 0.5 million jets) is utilised to monitor the training progress and avoid overfitting.

The training proceeds with a batch size of 256 jets, which balances computational efficiency and memory utilisation. The network weights are initialised using the He uniform initialiser [141]. The initial learning rate is set to 0.001, which is a common choice for the Adam optimiser [92] used to update the trainable weights and biases. Each layer uses the ELU (Exponential Linear Unit) activation function which has the form:

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha(\exp(x) - 1) & \text{if } x \leq 0 \end{cases} \tag{5.12}$$

The small but non-zero value for negative neuron outputs has been shown to help against the problem of vanishing gradients [142].

To ensure optimal model performance, an adaptive learning rate schedule is implemented with a decay factor of 0.03. This reduction is triggered when the validation loss fails to improve by at

least 0.1 over a patience period of 4 epochs. The training process incorporates an early stopping mechanism that terminates after 8 consecutive cycles of learning rate decay, indicating that further training would yield diminishing returns.

The complete training of the VAE required 33 epochs, with each epoch taking approximately one hour on the DEEPTHOUGHT GPU cluster hosted at the Institute of Experimental Particle Physics (ETP), Karlsruhe Institute of Technology (KIT), using a single `NVIDIA` Titan GPU with 12 GB of VRAM.

5.4. Data Control Region

As mentioned in Section 5.2.5, the search is performed in a pre-defined signal region. However, the VAE itself is trained on a Control Region (CR) dominated by the background of QCD multijet events, where the signal is comparatively suppressed. An effective CR should reduce the significance $N_{\text{sig}}/\sqrt{N_{\text{background}}}$ by a factor of at least 10 relative to the signal region, while maintaining sufficient statistical power. Furthermore, this region must be complementary and orthogonal to the signal region chosen for the search.

The complementarity criterion is satisfied by inverting the pseudorapidity separation condition that defines the analysis region. While the analysis region uses $|\Delta\eta| < 1.3$, a simple inversion to $|\Delta\eta| > 1.3$ is observed to be insufficient for signal suppression, as the fraction of events in this region from one specific signal model, this being the $W_{\text{KK}} \rightarrow \text{WR} \rightarrow \text{WWW}$ decay mentioned in Table 5.1, remains comparable to that in the analysis region. Therefore, a more stringent lower boundary of $|\Delta\eta| > 2.0$ is used to define the CR.

Additionally, to better replicate the m_{jj} distribution characteristics of the analysis region, an upper boundary of $|\Delta\eta| < 2.5$ is established. This constraint is physically motivated by the relation between dijet invariant mass and pseudorapidity separation in the highly relativistic limit:

$$m_{JJ}^2 = 2p_{\text{T},1}p_{\text{T},2}(\cosh\Delta\eta - \cos\Delta\phi), \tag{5.13}$$

where higher values of $|\Delta\eta|$ correlate with higher m_{jj}. The resulting CR, defined by $2.0 < |\Delta\eta| < 2.5$, yields an m_{jj} distribution that closely matches that of the signal region, with a slight increase in event count at lower m_{jj} values.

The W_{KK} signal is further suppressed by implementing additional criteria, as shown in Equation 5.14:

$$\text{AND} \begin{cases} 2.0 < |\Delta\eta| < 2.5 \\ \text{No third jet with } p_{\text{T}} > 300\,\text{GeV} \\ \text{OR} \begin{cases} \frac{|p_{\text{T},1} - p_{\text{T},2}|}{p_{\text{T},1} + p_{\text{T},2}} > 0.1 \\ A = p_{\text{T},1}p_{\text{T},2}(2\cosh\Delta\eta + 2)/m_{JJ}^2 \notin [0.95, 1]. \end{cases} \end{cases} \tag{5.14}$$

These additional requirements on jet p_{T} asymmetry further enhance the background enrichment in the control region. The detailed motivation and optimisation of these cuts are not elaborated further in this work.

5.4.1. Creating the Training Dataset for the VAE

Training the VAE on a dataset that accurately represents the background distribution in the signal region (SR) is crucial for anomaly detection. However, training the VAE itself is not as simple as using the data control region. Despite the fine-tuned and physically motivated cuts that are used to construct the CR, its kinematics are significantly different from that of the SR, and doing so would lead to a network strongly biased to these kinematics.

The kinematic distributions in the CR differ from the SR particularly in terms of the pseudorapidity separation between jets, which introduces biases in the jet p_{T} distributions and angular correlations. Training directly on CR data would create a model that considers typical SR kinematics as anomalous, defeating the purpose of the anomaly detection framework.

To overcome this challenge, a novel approach is implemented that maintains the background-only nature of the CR while approximating the kinematic distributions of the SR. This technique selectively samples jets from the CR that closely match the kinematic properties of jets in the SR, creating a hybrid dataset with SR-like kinematics but without signal contamination. The method employs a nearest-neighbor search using multidimensional binary search trees or k-d trees [143] in a three-dimensional feature space comprising the logarithm of the transverse momentum and the relative pseudorapidity and azimuthal angles of the jets.

The procedure is formalised in Algorithm 1, which details the construction of the training dataset through this kinematic matching process, while the resultant kinematics of the constructed dataset are shown in Figure 5.3.

Algorithm 1 VAE Input Dataset Construction

1: **Input:** Signal Region jets $\mathcal{J}_{\mathrm{SR}}$, Control Region jets $\mathcal{J}_{\mathrm{CR}}$
2: **Output:** dataset $\mathcal{J}_{\mathrm{TRAIN}}$ with SR kinematics and CR jets
3: initialise empty dataset $\mathcal{J}_{\mathrm{TRAIN}}$
4: **for** each jet J in $\mathcal{J}_{\mathrm{SR}}$ **do**
5: Compute feature vector $F_J = (log(p_{\mathrm{T}}(J)),\ \eta_{rel}(J),\ \phi_{rel}(J))$
6: Find nearest neighbor K_{nearest} in the k-d tree using the F_J vector
7: Add K_{nearest} to $\mathcal{J}_{\mathrm{TRAIN}}$
8: **end for**
9: **return** $\mathcal{J}_{\mathrm{TRAIN}}$

This approach ensures that the VAE is trained on a dataset that lacks potential signals while maintaining the kinematic distributions characteristic of the SR. By preserving the SR kinematics in the training data, the VAE becomes sensitive to genuine anomalies when applied to the actual SR, rather than being biased by the differences between the CR and SR kinematic distributions.

5.5. Reconstruction and Anomaly Detection Performance of the VAE

The VAE architecture represents an essential component of the overall anomaly detection framework. A thorough understanding of its reconstruction capabilities for both the SM

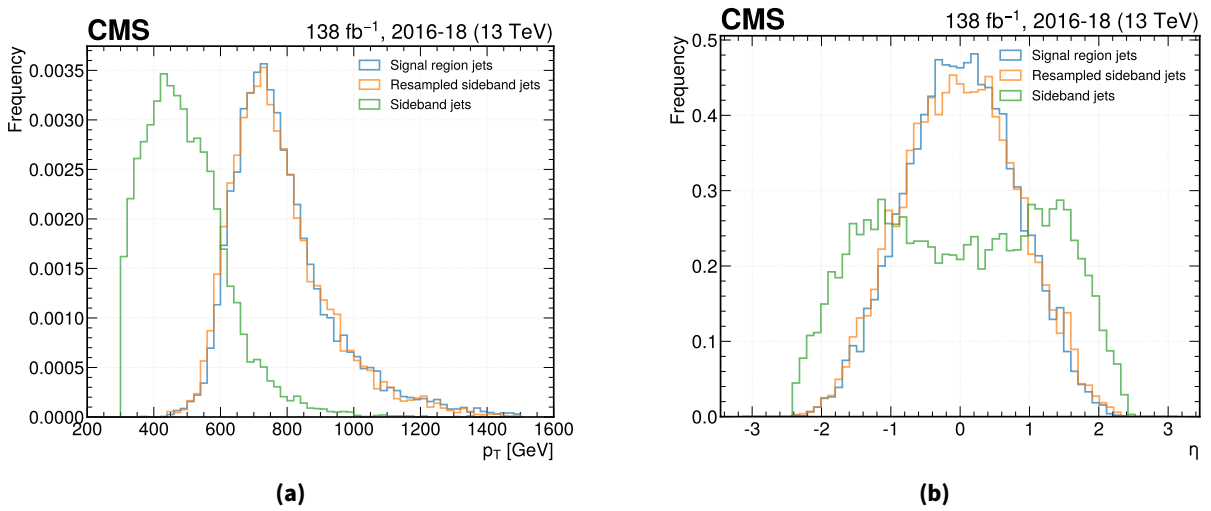


Figure 5.3.: p_T (left) and η (right) distributions of the leading jet J_1 for $\mathcal{J}_{CR}$, $\mathcal{J}_{SR}$ and $\mathcal{J}_{TRAIN}$

background and potential signal events is crucial for establishing its efficacy as an anomaly detection tool. This section presents a comprehensive analysis of the VAE's performance in reconstructing jet constituents and its ability to discriminate between background and signal events based on reconstruction loss alone. At the same time, it will be made clear why the usage of the loss alone does not suffice for this analysis.

The reconstruction performance of the VAE is evaluated and shown in the form of correlation matrix plots in Figure 5.4 for SR data events, in Figure 5.5 for the $W' \rightarrow B't$ decay with $M_{W'} = 3\,\text{TeV}$, $M_{B'} = 400\,\text{GeV}$, and in Figure 5.6 for the $X \rightarrow YY'$ process with $M_X = 3\,\text{TeV}$, $M_Y = M_{Y'} = 170\,\text{GeV}$. Each plot shows the correlation between input and reconstructed kinematic features, these being the momentum vector components $(p_T, \eta_{rel}, \phi_{rel})$ for the hardest, 4^{th} and 7^{th} hardest jet constituents.

A clear pattern emerges from these correlation matrices. While the VAE successfully reconstructs the hardest particle with high fidelity across all event types, there is a progressive deterioration in reconstruction quality for softer constituents. This degradation is more pronounced for signal events compared to the SR Data. As softer particles often carry distinctive information about multi-prong decay structures in signal events, their reduced reconstruction quality manifests in higher values of the event loss defined in Equation 5.11 for signal compared to SR Data events. This difference in reconstruction capability forms the basis for employing the VAE loss as an anomaly detection metric.

The distribution of VAE loss values for a set of benchmark signal models is presented in Figure 5.7. A clear trend is observed: signal models characterised by higher numbers of prongs consistently exhibit higher loss values. This aligns with physical intuition, as events with more complex decay topologies present greater challenges for the VAE to encode within its latent space, especially when trained predominantly on the simpler topologies of QCD jets.

To investigate the effect of decay kinematics on detection sensitivity, Figure 5.8 displays the loss distributions for a specific signal model where the mass of the parent particle is fixed while the masses of the daughter particles are varied. The results demonstrate that heavier daughter particles lead to higher loss values and consequently enhanced discrimination power. In contrast, lighter decay products generate events with kinematic features more closely resembling those of QCD jets, resulting in lower loss values. This behaviour confirms that the

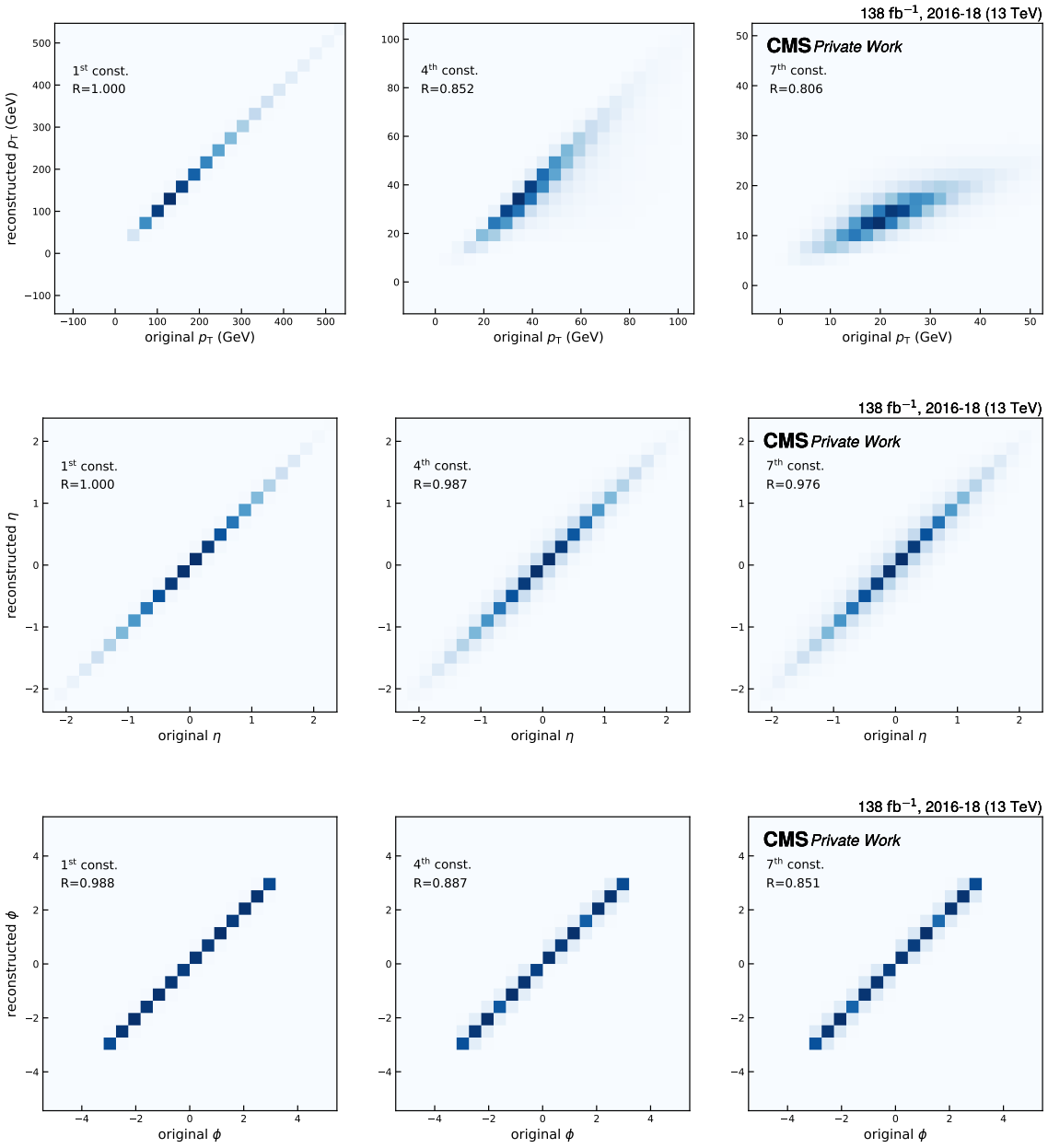


Figure 5.4.: Reconstruction Performance of the VAE on signal region jets, arranged by feature and shown for 3 representative jet constituent particles, with the Pearson Correlation Coefficient R being shown in the inset.

VAE, trained on raw particle momenta, exhibits sensitivity to the characteristic mass scales present in an event.

For a more granular analysis, Figure 5.9 presents the individual jet loss values per event for selected benchmark signals, where the two daughter decay products are deliberately chosen to have different masses. The results clearly demonstrate that jets containing the heavier decay product consistently exhibit larger loss values compared to those containing lighter particles. This observation highlights the VAE's sensitivity to the mass scale of the individual jets. However, to ensure robust anomaly detection and to avoid missing events where only one jet exhibits highly anomalous features, the event-level loss definition previously established in Equation 5.11 is employed. This conservative approach ensures that events with at least one

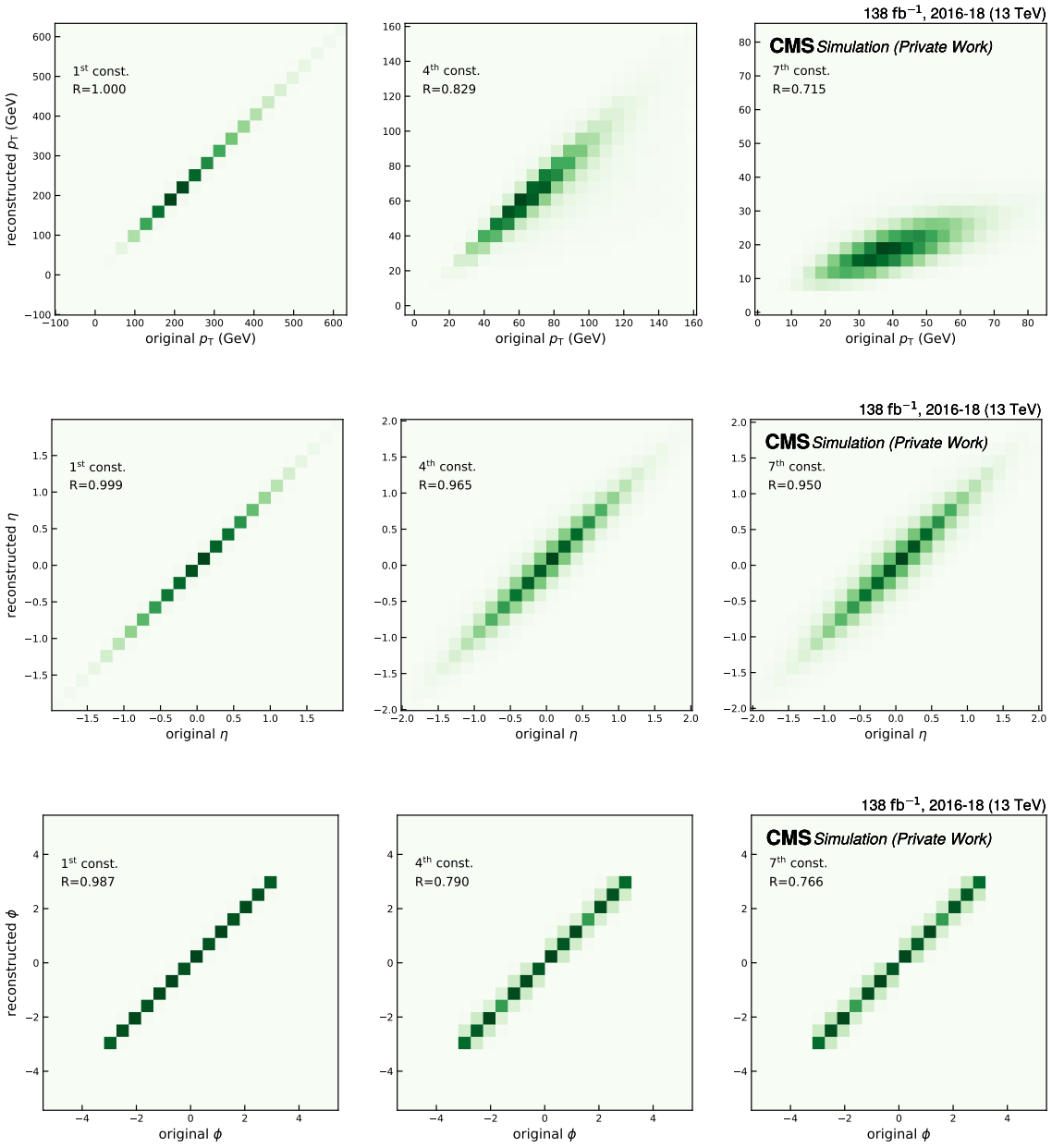


Figure 5.5.: Reconstruction Performance of the VAE on the simulated signal resonance $W' \rightarrow tB'$, arranged by feature and shown for 3 representative jet constituent particles, with the Pearson Correlation Coefficient R being shown in the inset.

significantly anomalous jet are appropriately flagged, maximizing sensitivity to potential new physics signatures.

The discrimination capability of the VAE loss as an anomaly metric is quantified through the ROC curves shown in Figures 5.10 and 5.11 for the previously discussed signal benchmarks. The corresponding AUC values generally indicate strong discrimination power, following the trends already observed in the loss distributions. Notably, although not presented here, similar enhancements in performance were also observed for scenarios with higher parent particle masses, consistent with expectations based on the increasing kinematic differences from the overwhelming QCD background that one expects to dominate the SR Data as well.

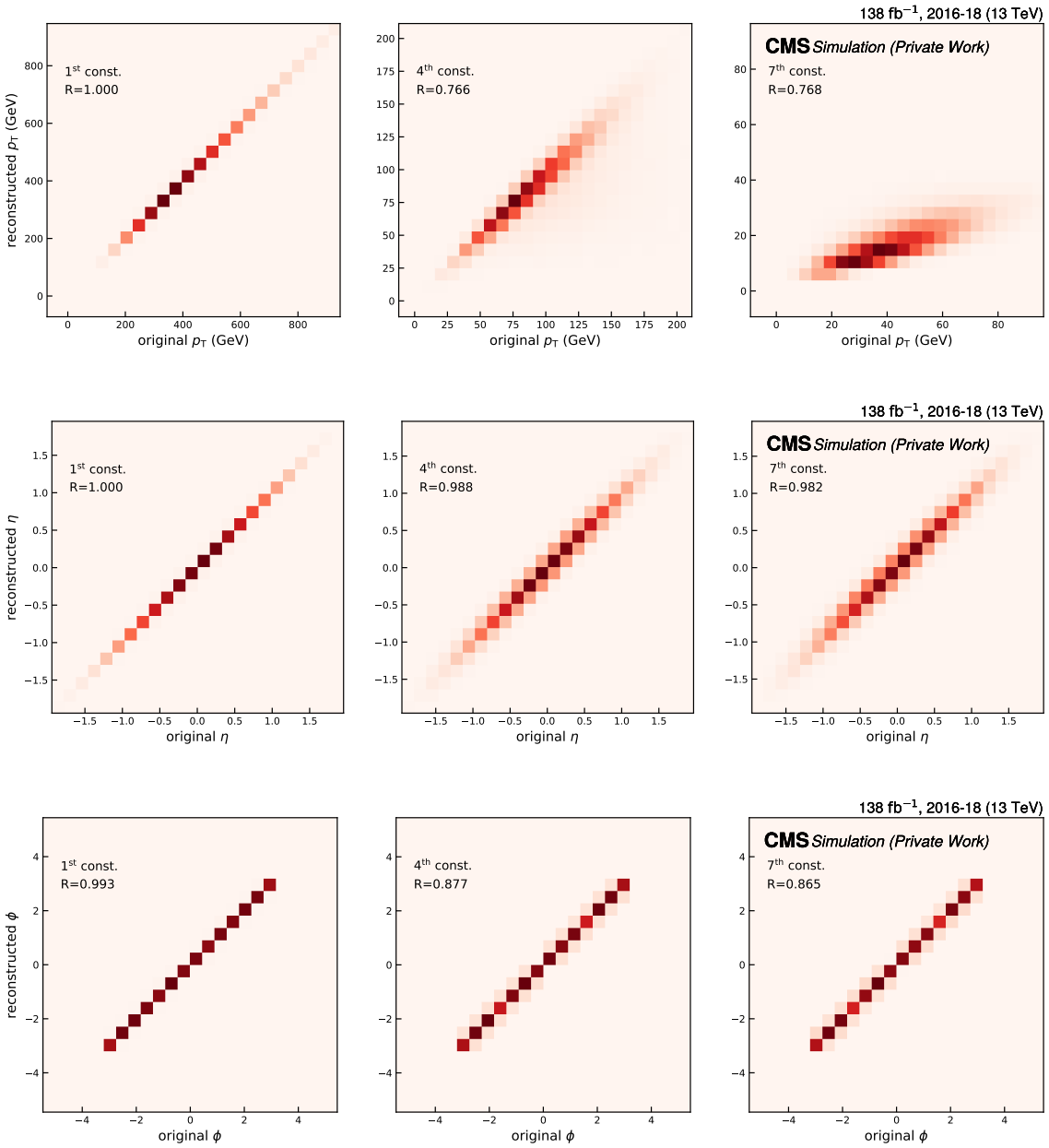


Figure 5.6.: Reconstruction Performance of the VAE on the simulated signal resonance $X \longrightarrow YY' \longrightarrow 4q$, arranged by feature and shown for 3 representative jet constituent particles, with the Pearson Correlation Coefficient R being shown in the inset.

5.5.1. Limitations of the VAE Loss and the Need for QR

Despite the promising discrimination capabilities demonstrated by the VAE loss, a critical limitation emerges when considering its practical application as an anomaly detection metric. The fundamental issue lies in the VAE's intrinsic sensitivity to the characteristic mass scales present in jets, which introduces an undesirable correlation between the reconstruction loss and the dijet invariant mass m_{jj}. This correlation is clearly illustrated in Figure 5.12, which presents a scatter plot of event-level VAE loss against m_{jj} for events in the SR Data. A substantial correlation is observed, quantified by a Pearson correlation coefficient of $R = 0.64$. This strong mass-loss relationship poses significant challenges for anomaly detection in a model-agnostic

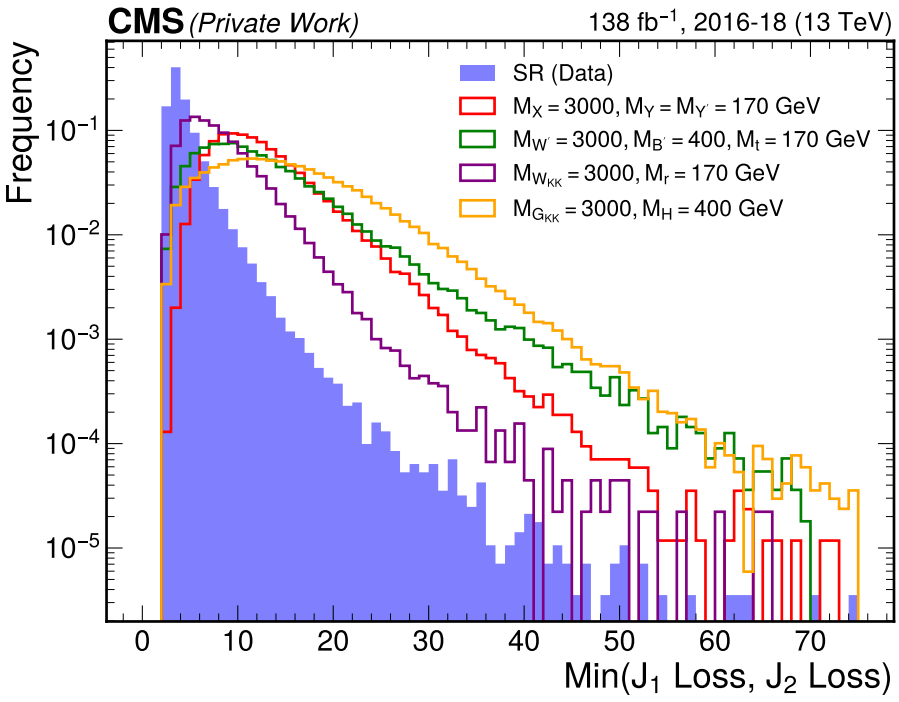


Figure 5.7.: Event-level losses for a set of benchmark signal models and SR Data

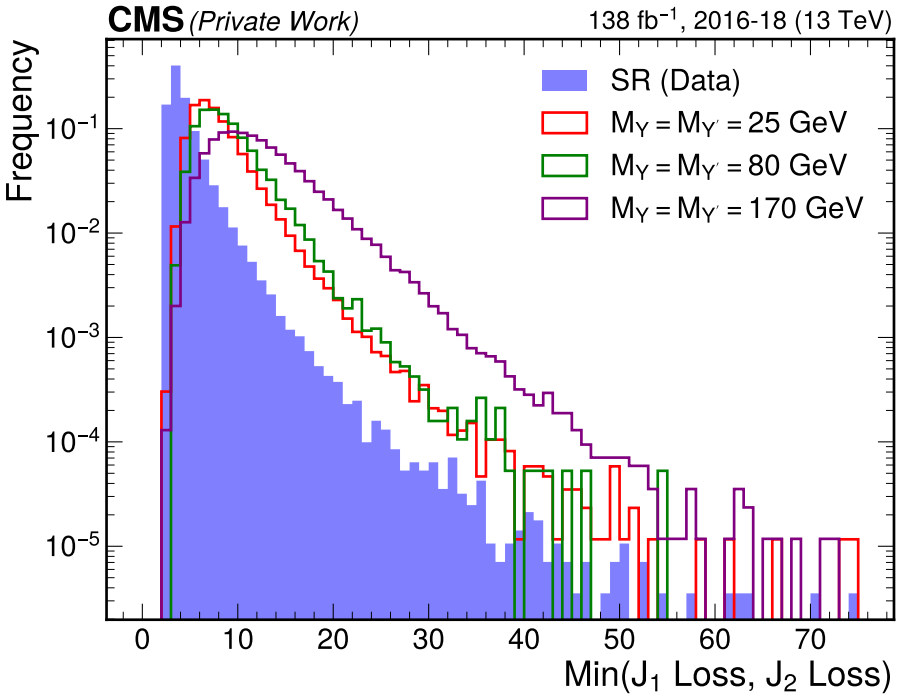


Figure 5.8.: Event-level losses for the benchmark signal model $X \rightarrow YY'$ and SR Data, with the mass of the parent resonance being fixed at $M_X = 3$ TeV and that of daughter decay products M_Y and $M_{Y'}$ being varied

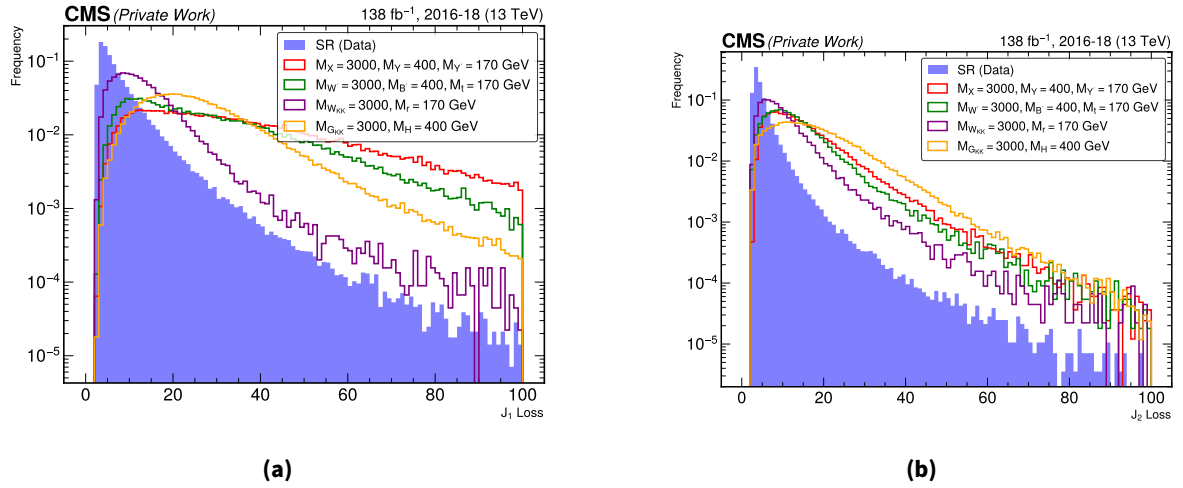


Figure 5.9.: Loss distributions for the leading jet J_1 and subleading jet J_2 for the SR Data and a set of benchmark signal models

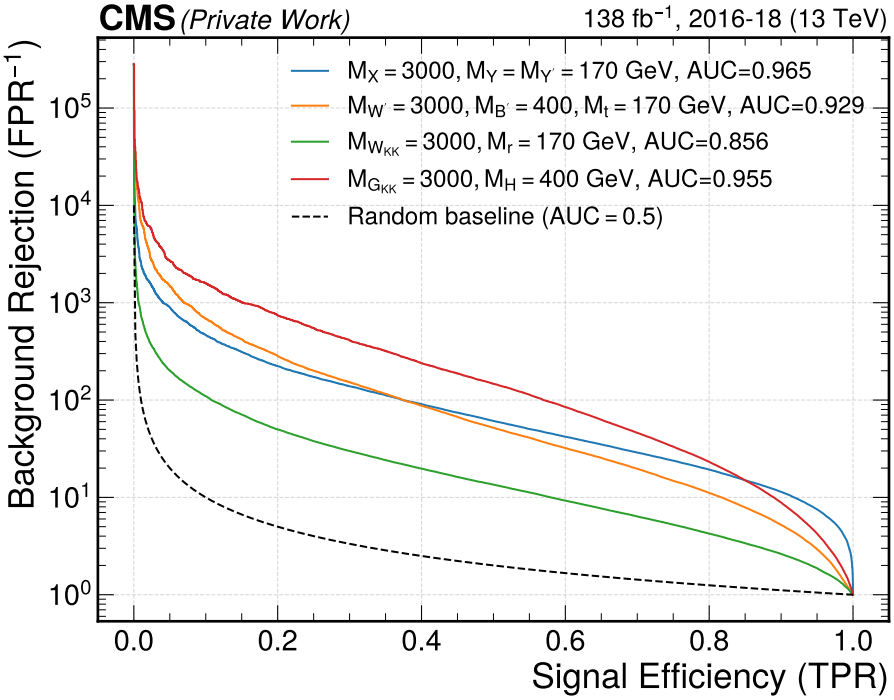


Figure 5.10.: ROC Curves with the corresponding AUC scores for a set of benchmark signal models and the SR Data

context. If one were to apply a simple threshold to the VAE loss for selecting a potentially signal-enriched sample of events, there would be a systematic bias toward preferentially selecting high-mass QCD events rather than genuine anomalies. This selection bias introduces a problematic phenomenon known as **mass sculpting**, where the resulting background m_{jj} distribution no longer exhibits the expected smoothly falling behaviour characteristic of QCD processes. Instead, the background distribution begins to artificially peak at certain mass values, potentially mimicking the shape expected from resonant decays that are characteristic of signal processes. Such sculpting significantly compromises the statistical power of bump-hunting

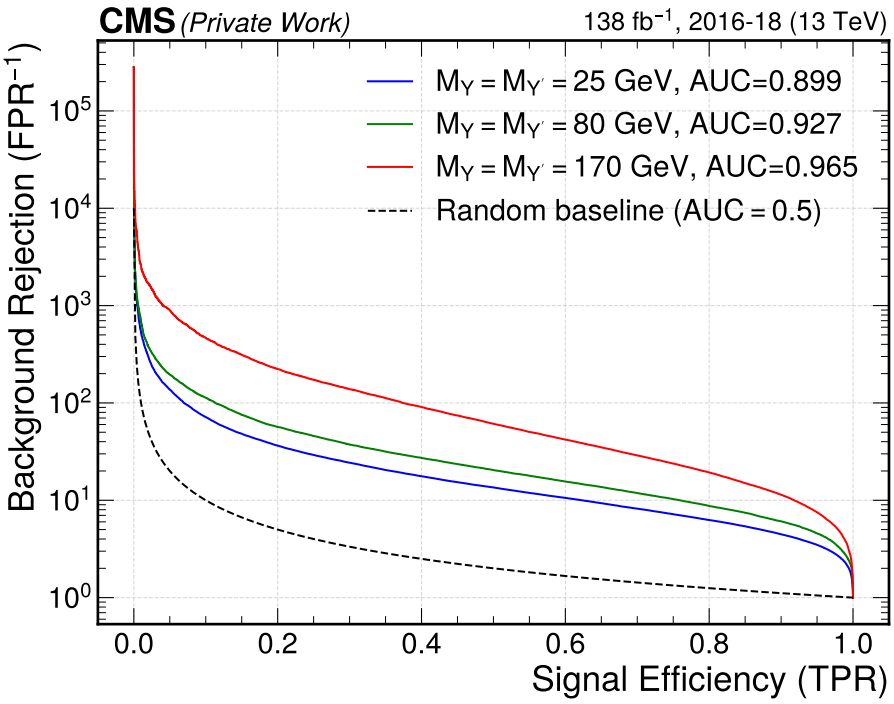


Figure 5.11.: ROC Curves with the corresponding AUC scores for the benchmark signal model $X \rightarrow YY'$ and SR Data, with the mass of the parent resonance being fixed at $M_X = 3$ TeV and that of daughter decay products M_Y and $M_{Y'}$ being varied

procedures that rely on the contrast between a smoothly falling background and a localised signal excess. To mitigate this mass correlation while preserving the anomaly detection sensitivity of the VAE, a more sophisticated approach becomes necessary. This requirement motivates the introduction of the QR technique as the subsequent step in the algorithm, a key part of the combined VAE-QR methodology presented in this thesis. The QR procedure effectively decorrelates the loss from the mass, enabling a truly model-agnostic search capable of effectively isolating anomalies across the entire m_{jj} spectrum without introducing artificial features in the background distribution.

5.6. Building VAE-QR: QR for Decorrelation

In high-energy physics searches, analyses are typically conducted across the spectrum of a variable of interest, which in this case is the dijet invariant mass m_{jj}. For the null hypothesis, representing the SM background expectation, this variable is expected to exhibit a specific, smoothly falling distribution. The fundamental objective is to identify statistically significant deviations from this expected shape, which could indicate the presence of new physics. Consequently, it is crucial that any selection criteria applied to the data do not artificially sculpt the distribution of m_{jj} or other variables of interest, as such sculpting would compromise the statistical integrity of the search and potentially lead to false signal identifications.

The application of a flat cut on the VAE loss, as illustrated in Figure 5.13, fails to satisfy this critical requirement. When one looks at Figure 5.14, it becomes evident that different event

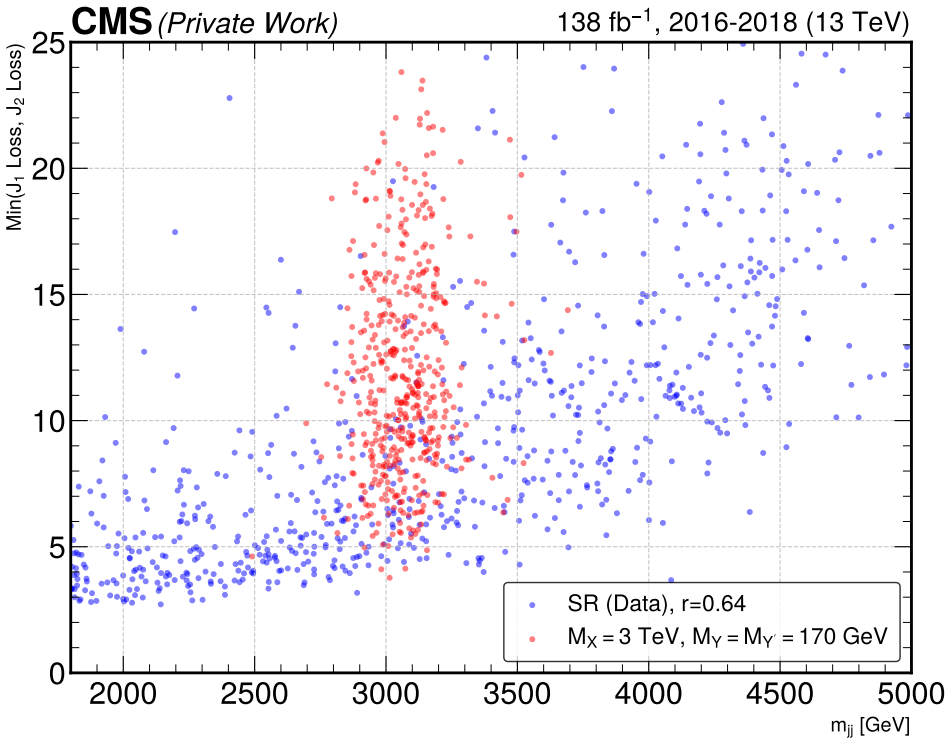


Figure 5.12.: Variation of the event-level VAE loss with the dijet invariant mass m_{jj}, showing an undesirable correlation between the two for the SR, which is expected to be overwhelmingly dominated by QCD multijet events

selection categories based on flat loss thresholds produce markedly different m_{jj} distributions. This sculpting effect is particularly problematic in regions where the background distribution begins to artificially mimic the expected shape of a resonant signal, thereby reducing sensitivity to genuine new physics and potentially introducing false bumps.

Various techniques have been developed in the literature to address the challenge of decorrelating discriminating variables from the mass spectrum. [144, 145]. For this analysis, the QR technique is employed, following the approach first formalised by Koenker and Bassett [146] and subsequently adapted for HEP applications.

The fundamental objective of the QR approach is straightforward yet powerful: rather than applying a uniform threshold to the event-level loss, the cut value is allowed to vary as a function of m_{jj}. This adaptive threshold accounts for the intrinsic correlation between the loss and mass distributions, enabling a selection that preserves the natural shape of the background m_{jj} spectrum.

The QR methodology proceeds by defining selection categories denoted hereafter as Q_X, where X represents a specific percentile. The goal is to learn a function $F_X = \mathcal{L}_T(m_{jj})$ such that for any particular value of m_{jj}, exactly $X\%$ of the events will have a loss greater than $\mathcal{L}_T(m_{jj})$, while the remaining $(100 - X)\%$ will have a loss below this threshold. It is important to emphasise that $\mathcal{L}_T(m_{jj})$ is not a fixed value but rather a continuous function that varies with m_{jj}, adaptively compensating for the mass-loss correlation observed in Figure 5.12.

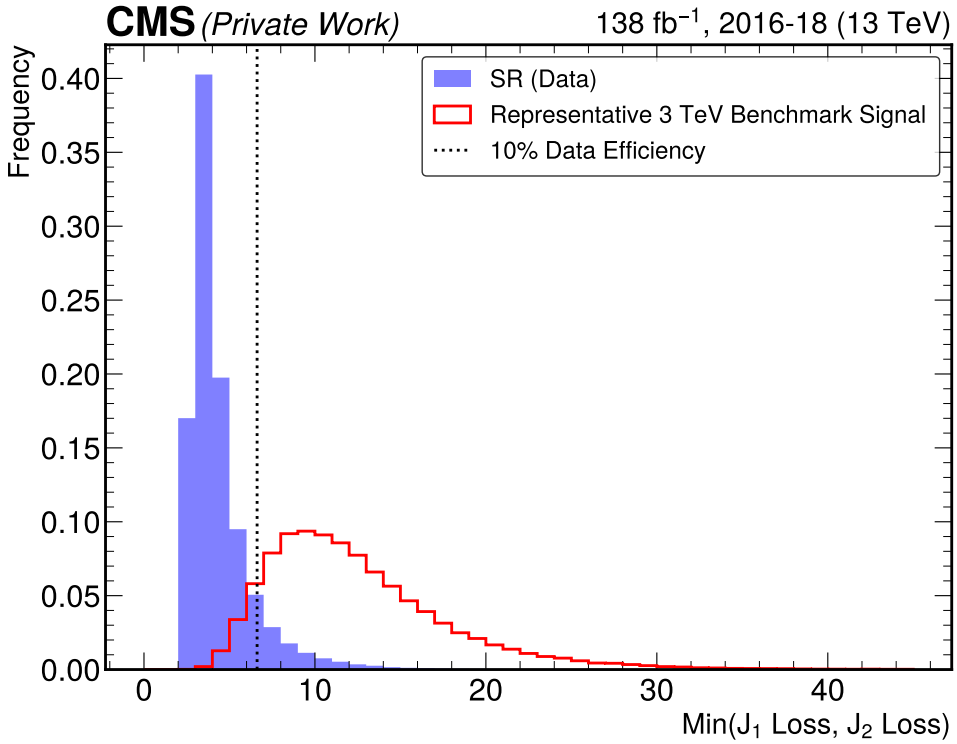


Figure 5.13.: Using a flat cut to select a specific percentage of events from the SR is shown to have undesirable effects

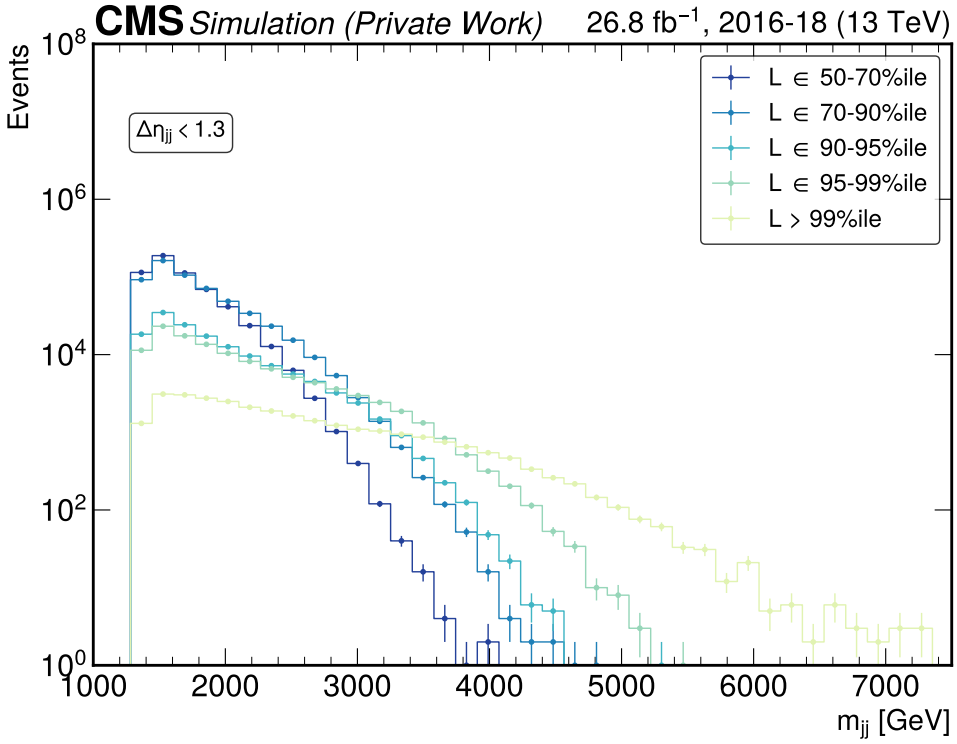


Figure 5.14.: Sculpted m_{jj} distributions across various categories when selected using flat cuts

This approach effectively provides a mechanism to select the $X\%$ most anomalous events from the dataset while accounting for the natural tendency of higher m_{jj} events to exhibit higher loss values due to their relative rarity in the QCD multijet background. The power of this technique lies in its ability to maintain consistent background shapes across different selection categories. For instance, selecting events between the 50^{th} and 70^{th} loss percentiles and comparing them to the 1% most anomalous events will yield identically shaped background distributions that differ only by a scaling factor proportional to the relative number of events in each category. Throughout this thesis, the term *category* is used to refer to each distinct selection defined using the QR percentiles.

Having established the conceptual framework of the QR approach for mass decorrelation, the subsequent sections will detail the technical implementation and performance of the QR within the context of the VAE-QR used in this analysis.

5.6.1. Mathematical Framework of the QR

Fundamentally, QR aims to determine a threshold function that, for a given m_{jj} value, predicts the loss value above which a specific fraction of events lies. In this analysis, the QR is implemented using a DNN architecture, with the technical implementation details elaborated in Section 5.6.2.

To properly contextualise QR, it is instructive to first consider more familiar regression techniques. In standard linear regression, the objective is to estimate the conditional mean of a response variable Y given predictor variables X, typically by minimising the Mean Squared Error (MSE):

$$\mathcal{L}_{\text{MSE}}(\hat{Y}, Y) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{5.15}$$

where Y_i represents the actual values and $\hat{Y}_i$ the predicted values. While the MSE loss is mathematically convenient and leads to an estimator of the conditional mean, it is sensitive to outliers due to the squared term.

A more robust alternative is median regression, which estimates the conditional median by minimising the Mean Absolute Error (MAE):

$$\mathcal{L}_{\text{MAE}}(\hat{Y}, Y) = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i| \tag{5.16}$$

The MAE loss leads to an estimator that corresponds to the 50^{th} percentile of the conditional distribution $P(Y|X)$. The natural extension to this is the QR, which enables one to find an arbitrary percentile instead.

The loss function that enables this is the asymmetric pinball or quantile loss [146], defined for a specified quantile $q \in (0, 1)$ as:

$$\mathcal{L}_q(Y, \hat{Y}) = \begin{cases} q \cdot (Y - \hat{Y}) & \text{if } Y \geq \hat{Y} \\ (1 - q) \cdot (\hat{Y} - Y) & \text{if } Y < \hat{Y} \end{cases} \tag{5.17}$$

This can be expressed more compactly as:

$$\mathcal{L}_q(Y, \hat{Y}) = \max\{q(Y - \hat{Y}), (q - 1)(Y - \hat{Y})\} \tag{5.18}$$

It is trivial to see that substituting $q = 0.5$ recovers the MAE loss.

$$\mathcal{L}_{q=0.5}(Y, \hat{Y}) = 0.5 \cdot |Y - \hat{Y}| \tag{5.19}$$

The pinball loss introduces an asymmetric penalty: predictions below the actual value incur a penalty weighted by q, while predictions above the actual value incur a penalty weighted by $(1 - q)$.

In the context of neural network training, minimising the pinball loss for a specific quantile q leads to a model whose predictions approximate the q-th quantile of the conditional distribution [147].

The complete QR framework in this analysis operates as follows: Given the dijet mass m_{jj} as input, the model is trained to predict the loss threshold $\mathcal{L}_T(m_{\text{jj}})$ such that a fraction q of events at that mass have a loss greater than $\mathcal{L}_T(m_{\text{jj}})$. The objective function for training is the pinball loss from Equation 5.17, which ensures that the learned threshold function properly estimates the desired quantile. This approach effectively learns the inverse of the conditional cumulative distribution function of the loss given the mass.

For the implementation in this analysis, a DNN architecture is employed to learn six different quantile thresholds simultaneously. While one could train six independent networks for each desired quantile, it is computationally more efficient to learn all thresholds concurrently through a multi-output network with a six-dimensional vector pinball loss, where each component corresponds to the loss function for a specific quantile threshold. This network $\mathcal{L}_T(m_{\text{jj}})$, accepts the m_{jj} as input and predicts the corresponding loss threshold values for all target quantiles in a single forward pass. This approach enables the partitioning of the signal region data into seven distinct categories: events below the lowest threshold, events between consecutive thresholds, and events above the highest threshold. In practice, as will be seen, three of these categories are used. By construction, each resulting category exhibits the same shape for the m_{jj} distribution, differing only in the relative number of events contained within each category. This methodology successfully achieves the desired decorrelation between the VAE loss and the mass spectrum. The partitioning strategy facilitates a statistically robust search for resonant anomalies while maintaining sensitivity across the entire m_{jj} range, as the background shape remains consistent regardless of which category is examined.

5.6.2. Technical Implementation

The QR component of the VAE-QR framework is implemented using the TensorFlow [139] package, providing a flexible computational framework for constructing and training the neural network architecture. Given the relationship between the VAE loss and m_{jj} to be learned is relatively simple but potentially nonlinear, a moderately complex fully connected neural network architecture is employed.

The network accepts a single input node corresponding to the dijet invariant mass m_{jj} of each event. To facilitate efficient training and numerical stability, the input values are normalised to have zero mean and unit standard deviation across the training dataset, as is standard practice.

The network architecture consists of five fully connected hidden layers, each containing 30 nodes. Each hidden layer uses the Exponential Linear Unit (ELU) activation function of Equation 5.12, which enables the modeling of nonlinearities while simultaneously mitigating the vanishing gradient problem that can affect deeper networks.

The output layer of the network consists of six nodes, each corresponding to one of the desired quantile thresholds: $q = [0.3, 0.5, 0.7, 0.9, 0.95, 0.99]$.

Central to the training of the QR network is the vectorised pinball loss function. For a set of quantiles $\mathbf{q} = [q_1, q_2, \ldots, q_6]$ and corresponding predictions $\hat{\mathbf{Y}} = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_6]$ for a true loss (this being the event-level VAE loss, to avoid ambiguity) value Y, the vectorised pinball loss is defined as:

$$\mathcal{L}_{\text{pinball}}(Y, \hat{\mathbf{Y}}, \mathbf{q}) = \sum_{i=1}^{6} \begin{cases} q_i \cdot (Y - \hat{Y}_i) & \text{if } Y \geq \hat{Y}_i \\ (1 - q_i) \cdot (\hat{Y}_i - Y) & \text{if } Y < \hat{Y}_i \end{cases} \tag{5.20}$$

Each component of this loss function corresponds to one predicted threshold, and minimising this loss ensures that approximately q_i fraction of the events will have VAE loss values above the predicted threshold $\hat{y}_i$ at any given m_{jj} value.

The network is trained using the ADAM optimiser [92]. The initial learning rate is set to 0.001, providing a good balance between convergence speed and stability. All network weights are initialised using the He initialiser [141].

To prevent overfitting and ensure optimal generalisation performance, a learning rate decay strategy is implemented. The learning rate is reduced by a factor of 0.2 with a patience of three epochs whenever the validation loss fails to improve. The training process is terminated through an early stopping mechanism after ten consecutive rounds of learning rate decay without improvement in the validation loss, indicating convergence of the optimisation process.

A critical consideration in the implementation of the QR is the efficient utilisation of the available SR data. Traditional model training would require partitioning the dataset into separate training, validation, and test subsets, potentially reducing the statistical power of the analysis. To maximise data utilisation while maintaining the statistical integrity of the model evaluation, a k-fold cross-validation approach is adopted, as detailed in the following subsection.

5.6.2.1. k-Fold Procedure for Training the QR

The k-fold cross-validation procedure employed in this analysis represents a sophisticated approach to maximise the statistical power of the available data while preventing any potential biases that might arise from using the same events for both training and evaluation. The entire SR dataset is divided into $k = 20$ equal parts, referred to as folds, ensuring that each fold contains a representative distribution of events across the entire m_{jj} spectrum.

On each fold, a separate QR DNN with the architecture described previously is trained, using an 80-20 split to create training and validation subsets, with the validation subset used for monitoring the training process, implementing learning rate decay, and triggering early stopping. This approach results in 20 independently trained QR networks, each specialised in predicting quantile thresholds for events that were not used in its training.

The critical aspect of this procedure is that for any event in a given fold F, the quantile thresholds are determined not by the network trained on fold F (which would introduce bias), but rather by the ensemble of networks trained on the remaining 19 folds. This ensures that each event is evaluated only by models that have never encountered it during training, maintaining the statistical independence necessary for unbiased threshold determination.

However, directly averaging the predictions from the 19 networks for each event could potentially introduce statistical fluctuations. To mitigate this effect and ensure smooth threshold functions, an intermediate smoothing step is incorporated into the procedure:

First, m_{jj} bins of unit width are defined across the range of interest (1460 GeV to 6800 GeV), covering the entire spectrum of dijet masses relevant to the analysis. For each bin center, the average prediction from the 19 networks (excluding the one trained on the fold containing the event) is calculated, providing a discretised mapping of m_{jj} values to quantile thresholds.

Next, to obtain a continuous, smooth threshold function, a third-order polynomial $P_3(m_{jj})$ is fitted to these discretised threshold values using the LMFIT[1] library. Crucially, the fitting procedure employs an event weighting scheme of the form $\exp(-(m_{jj} - 1455)/500)$, which assigns greater importance to the more statistically abundant low-mass bins. This weighting scheme ensures that the fitted polynomial accurately models the densely populated low-mass region while still providing reasonable extrapolation to the sparser high-mass region.

The resulting third-order polynomial for each fold provides a smoothed, continuous function that maps any m_{jj} value to the corresponding quantile thresholds. These polynomial functions are then used to categorise each event in the dataset based on its m_{jj} value and VAE loss, placing it into one of the six categories defined by the quantile thresholds.

There are two critical considerations in the implementation of the QR:

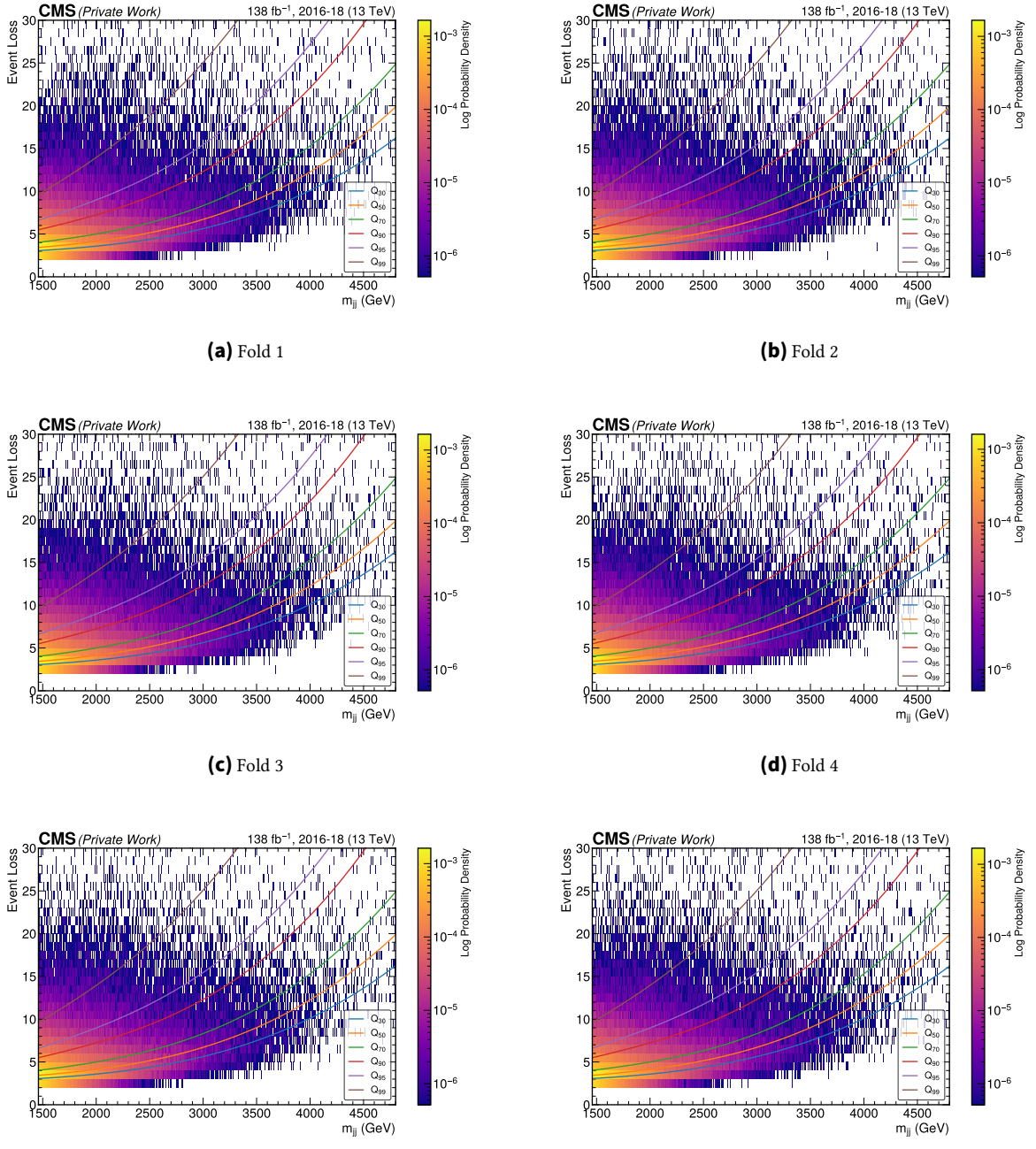
- efficient utilisation of the available SR data
- handling biases due to the presence of signal in the data

That being said, the QR is vulnerable to bias in the presence of large quantities of signal. Should such significant concentrations of a resonant signal exist around a certain m_{jj} value, these events would cluster in a particular region of the loss-m_{jj} phase space, potentially biasing the learned threshold functions. This potential bias can be mitigated to some extent through the

[1] https://lmfit.github.io/lmfit-py/

k-folding strategy, operating under the implicit assumption that each individual fold contains a negligible amount of any potential signal, these quantities being insufficient to substantially bias the QR DNN. The efficacy of this approach in preserving sensitivity while avoiding quantile function bias is rigorously validated through signal injection bias tests, as described in Appendix B.2.

The results of the k-fold splitting procedure and the QR training are shown in Figure 5.15 for the first six folds.



(a) Fold 1

(b) Fold 2

(c) Fold 3

(d) Fold 4

(e) Fold 5

(f) Fold 6

Figure 5.15.: Quantile regression polynomials fitted to background data for different folds. Each plot shows the distribution of event loss versus dijet mass (m_{jj}) with logarithmic density scaling. The coloured lines represent different quantile regression models (30%, 50%, 70%, 90%, 95%, and 99%) that characterise the threshold at various percentiles.

5.6.3. Summary

The two components of the VAE-QR methodology presented in this section together constitute a powerful approach for model-agnostic anomaly detection. Through the combination of a VAE for reconstruction of QCD background dominated jets, coupled with a QR for m_{jj} decorrelation, the technique successfully divides the SR data into several categories that, by construction, maintain identical smoothly falling background shapes in the absence of any signal. This decorrelation property is essential for the resonant anomaly search procedure that will be formally introduced in subsequent sections. As a demonstration of the efficacy of this approach, the background m_{jj} distributions for the mock MC QCD dataset, along with a representative 3 TeV resonant signal, are shown for the three most anomalous categories (corresponding to the highest VAE loss values) in Figure 5.16. The background spectra remain unsculpted across all categories within statistical uncertainties, confirming the successful implementation of the mass decorrelation strategy. Crucially, in the presence of signal, these categories will exhibit markedly different signal-to-background ratios, with the most anomalous categories showing substantial signal enrichment. This differentiation becomes particularly evident when performing the search over the m_{jj} spectrum, where the most anomalous categories demonstrate significantly enhanced sensitivity to potential new physics signatures while maintaining statistical robustness against false positive identifications.

5.7. Introduction to Statistics for HEP

Experimental results at the CMS experiment require a sophisticated statistical treatment to extract the parameters of interest while accounting for various sources of uncertainty. This section describes one of the most powerful tools, which is the binned profile likelihood approach used for parameter estimation and hypothesis testing.

5.7.1. Binned Profile Likelihood Approach

The standard practice within the CMS Collaboration is to employ a frequentist interpretation, with the likelihood function serving as a fundamental tool that quantifies the probability of observing the experimental data given a specific model with its associated parameters. In a typical analysis, the data consists of event counts distributed across multiple bins of an observable, which in the context of this analysis is the dijet invariant mass m_{jj}. The likelihood function for such binned data can be expressed as a product of Poisson probabilities [148]:

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_{i=1}^{n_{\text{bins}}} \text{Pois}(n_i | \nu_i(\mu, \boldsymbol{\theta})) = \prod_{i=1}^{n_{\text{bins}}} \frac{\nu_i(\mu, \boldsymbol{\theta})^{n_i}}{n_i!} e^{-\nu_i(\mu, \boldsymbol{\theta})}, \tag{5.21}$$

where n_i represents the observed number of events in bin i, and likewise $\nu_i(\mu, \boldsymbol{\theta})$ denotes the expected number of events, usually obtained from a background fit function. The parameter μ is the primary parameter of interest (this being the signal strength), while $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$ are the so-called *nuisance* parameters that account for various systematic uncertainties.

The expected event yield ν_i in each bin can typically be modeled as a sum of signal and background contributions:

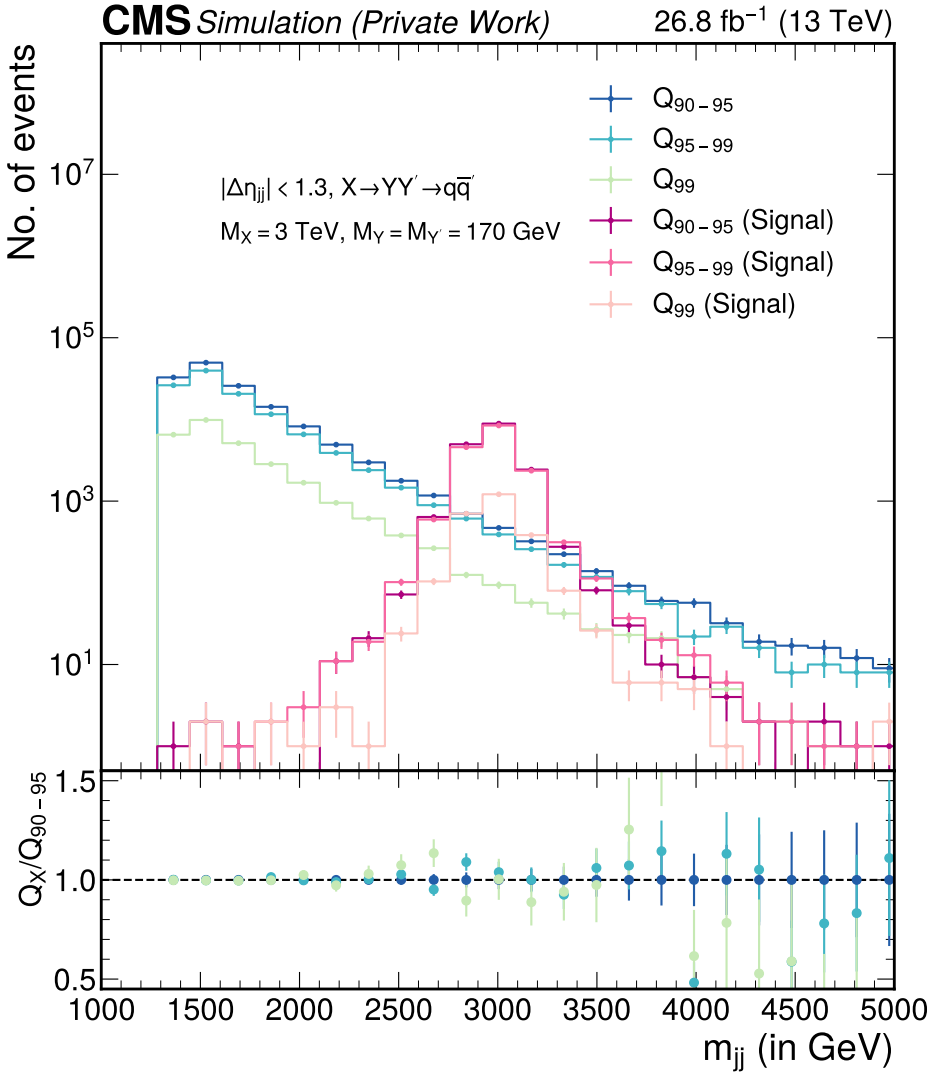


Figure 5.16.: Background and signal shapes for each of the three categories later used in the limit setting procedure, demonstrating the efficacy of the QR at constraining the background shape across categories, while simultaneously enhancing the signal fraction.

$$\nu_i(\mu, \boldsymbol{\theta}) = \mu \cdot s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}), \tag{5.22}$$

where $s_i(\boldsymbol{\theta})$ represents the expected signal yield and $b_i(\boldsymbol{\theta})$ represents the expected background yield in bin i. These expected yields are obtained from the respective template functions for background (see Equation 5.41) and signal (see Equation 5.44). Both of these components depend on the nuisance parameters $\boldsymbol{\theta}$.

5.7.2. Profile Likelihood Ratio Test Statistic

To assess the compatibility of the observed data with a specific value of μ, a test statistic based on the profile likelihood ratio is constructed [148]:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\boldsymbol{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})}, \tag{5.23}$$

where $\hat{\boldsymbol{\theta}}_\mu$ represents the set of nuisance parameter values that maximise the likelihood for a specific value of μ, and $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are the global maximum likelihood estimators that maximise $\mathcal{L}(\mu, \boldsymbol{\theta})$ without constraints. By construction, the profile likelihood ratio satisfies $0 \leq \lambda(\mu) \leq 1$, with values closer to 1 indicating better agreement between the data and the hypothesised value of μ.

It is often simpler to use the negative logarithm of the profile likelihood ratio as the test statistic, which is defined as:

$$t_\mu = -2 \ln \lambda(\mu) = -2 \ln \frac{\mathcal{L}(\mu, \hat{\boldsymbol{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})}. \tag{5.24}$$

This test statistic approximately follows a logarithmic χ^2 distribution with one degree of freedom in the asymptotic limit of large sample sizes, according to Wilks' theorem [149]. The best-fit value of the parameter of interest corresponds to $\hat{\mu}$, which minimises t_μ (equivalently, maximises $\lambda(\mu)$).

5.7.3. Treatment of Systematic Uncertainties

Systematic uncertainties quantify potential biases in the experimental setup, calibration, theoretical models, or simulation techniques that affect the interpretation of the data. Unlike statistical uncertainties, which decrease with increasing data, systematic uncertainties do not scale in a similar manner and require a careful treatment.

As mentioned earlier, these systematic uncertainties are incorporated through nuisance parameters $\boldsymbol{\theta}$, which modify the expected signal and background predictions according to the estimated effects of each uncertainty source. One way to constrain them is to use a set of auxiliary measurements $\tilde{\theta}_j$ that provide information about the nuisance parameters. The likelihood function can then be expressed as a product of Poisson probabilities for the observed data and a set of constraints on the nuisance parameters:

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_{i=1}^{n_{\text{bins}}} \text{Pois}(n_i | \nu_i(\mu, \boldsymbol{\theta})) \cdot \prod_{j=1}^{n_{\text{nuisance}}} C(\theta_j | \tilde{\theta}_j), \tag{5.25}$$

where $C(\theta_j | \tilde{\theta}_j)$ is a likelihood term that constrains the nuisance parameter θ_j based on auxiliary measurements.

5.7.4. Constraining Systematic Uncertainties

Nuisance parameters θ_j that describe systematic uncertainties are often not left unconstrained in the likelihood. Two main approaches exist to incorporate additional information about these parameters: using auxiliary measurements and incorporating prior knowledge.

An auxiliary measurement $\tilde{m}_j$ can be performed to obtain direct information about the true value of θ_j, where this measurement is statistically independent of the signal-region data [150]. These auxiliary measurements are often implemented as separate histograms (or control regions) in the fit. The global likelihood can be expressed as:

$$\mathcal{L}(\mu, \boldsymbol{\theta}) \;=\; \prod_{i=1}^{n_{\text{bins}}} \text{Pois}\big(n_i \mid \nu_i(\mu, \boldsymbol{\theta})\big) \;\times\; \prod_{j=1}^{m} \mathcal{L}_{\text{aux}}\big(\tilde{m}_j \mid \theta_j\big). \tag{5.26}$$

Another choice is to utilise pre-existing knowledge of the parameter θ_j, usually from other experiments, if this exists. Such knowledge can be incorporated using a functional prior $\pi(\theta_j)$. In a frequentist context, these terms can be interpreted as representing the distribution of an auxiliary measurement [151]. The likelihood, after the incorporation of such priors, becomes:

$$\mathcal{L}(\mu, \boldsymbol{\theta}) \;=\; \prod_{i=1}^{n_{\text{bins}}} \text{Pois}\big(n_i \mid \nu_i(\mu, \boldsymbol{\theta})\big) \;\times\; \prod_{j=1}^{m} \pi(\theta_j). \tag{5.27}$$

A common choice for $\pi(\theta_j)$ is a Gaussian prior:

$$\pi(\theta_j) \;=\; \frac{1}{\sqrt{2\pi}\,\sigma_j} \exp\left[-\frac{\big(\theta_j - \mu_j\big)^2}{2\sigma_j^2}\right], \tag{5.28}$$

where μ_j is the prior central value and σ_j represents the prior uncertainty on θ_j [152].

These constraint terms enter the fit alongside the main likelihood and are *profiled* out during maximum likelihood estimation [150, 148]. This involves finding the values of nuisance parameters that maximise the likelihood for each value of the parameter of interest using tools such as RooFit [153].

5.7.5. Signal Shape and Normalisation Uncertainties

Systematic uncertainties affecting the signal can broadly be categorized into two types: shape uncertainties and normalisation uncertainties.

Shape uncertainties affect the distribution of the signal across the bins of the discriminating variable without necessarily changing the total signal yield. These uncertainties might arise from imperfect modelling of detector resolution, energy scale calibrations, or theoretical uncertainties in differential distributions. Normalisation uncertainties, on the other hand, affect the overall signal yield without changing its shape. These uncertainties include those related to integrated luminosity, theoretical cross-section calculations, selection efficiencies,

and acceptance factors, and directly scale the expected signal yield across all bins by a common factor.

5.7.6. Log-Normal Treatment of Normalisation Uncertainties

While Gaussian constraints are commonly applied to nuisance parameters, normalisation uncertainties can be better modeled using log-normal distributions. This approach is appropriate for cases where the effect of the systematic variation must remain positive definite, such as efficiencies or scale factors [152].

In the log-normal approach, the expected yield is parameterised as:

$$\nu_i(\mu, \boldsymbol{\theta}) = \mu \cdot s_i \cdot \prod_j e^{\kappa_{ij}\theta_j} + b_i \cdot \prod_k e^{\eta_{ik}\theta_k}, \tag{5.29}$$

where κ_{ij} and η_{ik} represent the impact of nuisance parameter θ_j or θ_k on the signal or background prediction in bin i, respectively. The nuisance parameters θ_j and θ_k remain constrained by standard Gaussian terms, but their effect is exponential.

The log-normal parameterisation offers several advantages over a linear (Gaussian) treatment of normalisation uncertainties:

1. It ensures that predicted event yields remain positive.
2. It handles large uncertainties (e.g., greater than 20%) where the asymmetry of confidence intervals becomes important.

Mathematically, if a quantity X has a relative uncertainty of σ, a log-normal distribution with parameter $\kappa = \ln(1 + \sigma)$ provides the appropriate probability distribution. The corresponding nuisance parameter θ then modifies the nominal value X_0 as $X = X_0 \cdot e^{\kappa\theta}$, where θ follows a standard normal distribution.

For small uncertainties, the log-normal treatment approximates the Gaussian case, but it diverges significantly for larger uncertainties, providing more realistic confidence intervals. This approach has become standard in CMS analyses for handling normalisation uncertainties in a statistically robust manner [5].

5.7.7. Profiling Strategy and Statistical Interpretation

The complete fitting strategy involves finding the maximum likelihood estimates for both the parameter of interest μ and the nuisance parameters $\boldsymbol{\theta}$. This is achieved through numerical optimisation techniques that maximise the likelihood function (or, equivalently, minimise $-2\ln\mathcal{L}$).

The profiling procedure involves two steps:

1. Unconditional fit: Find the global maximum likelihood estimators $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ that maximise $\mathcal{L}(\mu, \boldsymbol{\theta})$ without constraints.

2. Conditional fits: For each tested value of μ, find the set of nuisance parameters $\hat{\boldsymbol{\theta}}_\mu$ that maximise $\mathcal{L}(\mu, \boldsymbol{\theta})$ with μ fixed.

The profiling approach effectively incorporates the impact of systematic uncertainties, allowing the nuisances to be constrained from the data. This results in a reduction of the overall uncertainty on the parameter of interest compared to a simplified approach where systematic uncertainties are treated as fixed [148]. Once the profile likelihood ratio test statistic $t_\mu = -2 \ln \lambda(\mu)$ is constructed, as defined in Eq. 5.24, its observed value $t_{\mu,\text{obs}}$ is determined from the actual experimental data. Thereafter a p-value can be calculated as the probability of obtaining a test statistic value at least as extreme as the observed one, assuming a specific hypothesis H_μ (corresponding to a signal strength μ):

$$p_\mu = \int_{t_{\mu,\text{obs}}}^{\infty} f(t_\mu|\mu)\, dt_\mu, \tag{5.30}$$

where $f(t_\mu|\mu)$ is the sample distribution of the test statistic under the hypothesis H_μ. This distribution of the test statistic is non-trivial to compute, but in the asymptotic limit of large sample sizes, it approaches a χ^2 distribution with one degree of freedom for the case of a single parameter of interest [149, 148].

The significance of an excess above the expected background is quantified by testing the background-only hypothesis ($\mu = 0$). The corresponding p-value p_0 represents the probability that background fluctuations alone would yield a result at least as signal-like as the one observed. This p-value is conventionally converted to a significance Z as follows:

$$Z = \Phi^{-1}(1 - p_0), \tag{5.31}$$

where Φ^{-1} is the inverse of the Cumulative Distribution Function (CDF) of the standard normal distribution. By convention in HEP, a significance of $Z \geq 5$ (corresponding to $p_0 \leq 2.87 \times 10^{-7}$) is required to claim a discovery [5]. This is the so-called 5σ criterion, making it very unlikely for the observed excess to arise from background fluctuations.

To evaluate the expected sensitivity of an analysis before examining the actual data, a special artificial dataset known as the *Asimov* dataset is employed [148]. In this dataset, all statistical fluctuations are ignored, and the observed counts in each bin exactly match the predicted counts for a specific model:

$$n_i^{\text{Asimov}} = \nu_i(\mu', \boldsymbol{\theta}'), \tag{5.32}$$

where μ' and $\boldsymbol{\theta}'$ represent the assumed values of the signal strength and nuisance parameters used to generate the Asimov dataset. Typically, $\mu' = 1$ is used to estimate the expected sensitivity for the signal model in question. The Asimov dataset is used to estimate the median expected test statistic distribution without performing extensive pseudo-experiments. This is most useful for complex analyses containing several nuisance parameters, as it allows for the expected sensitivities under such circumstances to be calculated with minimal effort [148].

Thereafter, one may wish to set upper limits on the signal strength parameter μ. For a specified confidence level (CL) of $1 - \alpha$ (typically 95%, corresponding to $\alpha = 0.05$), the upper limit μ_{up} is defined as the value for which:

$$p_{\mu_{\text{up}}} = \alpha, \tag{5.33}$$

where $p_{\mu_{\text{up}}}$ is the p-value for the signal-plus-background hypothesis with signal strength μ_{up}. In the asymptotic limit, this upper limit can be determined from the condition:

$$t_{\mu_{\text{up}}} = q_\alpha, \tag{5.34}$$

where q_α is the α-quantile of the test statistic distribution [148].

For setting upper limits, the CMS experiment employs a one-sided test statistic, defined as:

$$q_\mu = \begin{cases} -2 \ln \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}_\mu)}{\mathcal{L}(0, \hat{\hat{\boldsymbol{\theta}}}_0)} & \text{if } \hat{\mu} < 0 \\ -2 \ln \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} & \text{if } 0 \leq \hat{\mu} \leq \mu \\ 0 & \text{if } \hat{\mu} > \mu \end{cases} \tag{5.35}$$

This ensures that upward fluctuations of the data beyond the predicted signal strength do not count as evidence against the signal hypothesis [148].

This standard approach to setting upper limits can sometimes lead to exclusion of signal models to which the experiment has limited sensitivity, particularly when the observed number of events fluctuates even below the expected background. To address this issue, the CL_s technique [154], described in the following Section, was developed.

5.7.8. The CL_s Technique for Setting Upper Limits

The CL_s technique represents a *modified-frequentist* approach developed specifically to address certain limitations of standard frequentist methods when setting exclusion limits in particle physics searches [154, 155].

5.7.8.1. Mathematical Formulation

In the CL_s framework, one uses the test statistic q_μ to construct:

$$\text{CL}_{s+b} = P(q_\mu \geq q_{\text{obs}} | \mu s + b), \tag{5.36}$$

which represents the probability, under the signal-plus-background hypothesis with signal strength μ, to observe a test statistic value at least as extreme as the observed value q_{obs}. Similarly,

$$\text{CL}_b = P(q_{\mu=0} \geq q_{\text{obs}} | b), \tag{5.37}$$

represents the probability, under the background-only hypothesis q_0, with $\mu = 0$, to observe a test statistic at least as extreme as the observed value.

The CL_s test statistic is then constructed as:

$$\mathrm{CL}_s = \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b}. \tag{5.38}$$

It is important to note that CL_{s+b} and CL_b are not p-values in the traditional sense of hypothesis testing. Instead, they represent the probabilities of observing test statistic values at least as signal-like as the observed value, under their respective hypotheses.

If now one defines p_μ as the p-value for testing the signal-plus-background hypothesis and p_b is the corresponding p-value for the background-only hypothesis, then:

$$\mathrm{CL}_s = \frac{p_\mu}{1 - p_b} = \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b}. \tag{5.39}$$

A signal model with strength μ is considered excluded at confidence level $1 - \alpha$ if $\mathrm{CL}_s \leq \alpha$, with the excluded limit being obtaining by solving $\mathrm{CL}_s = \alpha$. This is in contrast to the standard frequentist approach, which would exclude the model if $\mathrm{CL}_{s+b} \leq \alpha$.

The conservative nature of the CL_s method arises because of the normalisation factor CL_b in its denominator. If there exists a significant downward fluctuation relative to the background expectation, CL_b becomes small, leading to a larger CL_s value than would be obtained in the standard approach. This mechanism prevents the exclusion of signal models in cases where such exclusion would arise due to statistical fluctuations rather than any real sensitivity. Mathematically, since $\mathrm{CL}_b \leq 1$ by definition (as a probability), it follows that $\mathrm{CL}_s \geq \mathrm{CL}_{s+b}$. Consequently, any signal model excluded by the CL_s method would be excluded by the standard approach as well, but not the other way round.

When the signal and background distributions are nearly indistinguishable, both CL_{s+b} and CL_b test statistics are similar in value, leading to a CL_s ratio close to 1, preventing a false exclusion. Only when there is genuine discriminating power and the data truly does not fit the signal-plus-background hypothesis compared to the background-only case, does the CL_s value decrease to an extent that permits one to confidently claim exclusion [151, 150]. This *protection* is particularly important in searches for new physics.

In practical implementations, the `Combine` tool used by CMS employs an *asymptotic limit approximation* to compute CL_s values efficiently. This makes use of Wilks' theorem [149, 148], which states that the profile likelihood ratio test statistic q_μ asymptotically follows a χ^2 distribution under the null hypothesis. The asymptotic approximation speeds up the computation of CL_s by exploiting the statistical nature of the profile likelihood ratio test statistic $q_\mu = -2\ln\left(\mathcal{L}(\mu, \hat{\hat{\theta}}_\mu)/\mathcal{L}(\hat{\mu}, \hat{\theta})\right)$, where $\hat{\mu}$ and $\hat{\theta}$ are the unconditional maximum likelihood estimates, and $\hat{\hat{\theta}}_\mu$ are the nuisance parameters optimised keeping μ fixed at a given value. As Wilks' theorem states, in the asymptotic limit (large data samples and well-constrained systematic uncertainties), the distribution of q_μ is a χ^2 distribution with one degree of freedom. Consequently, the p-values CL_{s+b} and CL_b can be approximated using the cumulative distribution function (CDF) of the χ^2 distribution rather than requiring the computationally intensive generation of toy experiments. For example, $\mathrm{CL}_{s+b} \approx 1 - \Phi\left(\sqrt{q_\mu}\right)$, where Φ is the standard normal CDF, and $\sqrt{q_\mu}$ follows a half-normal distribution under the $\mu s + b$ hypothesis.

Similarly, under the background-only hypothesis ($\mu = 0$), the test statistic follows a χ^2 mixture distribution:

$$CL_b \;=\; 1 - \tfrac{1}{2}\,\mathrm{CDF}\big(\sqrt{q_0}\big) \;-\; \tfrac{1}{2} \tag{5.40}$$

but in the asymptotic regime, CL_b simplifies to a complementary error function. By using this approximation instead of generating simulated toy Monte Carlo datasets, computational requirements are reduced significantly. The approximation remains valid when likelihoods are well-behaved and statistical models satisfy regularity conditions. However, in cases with sparse data or non-Gaussian uncertainties, the full procedure still becomes necessary. This balance makes the asymptotic method particularly advantageous for highly data-intensive CMS analyses.

To conclude, the CL_s technique is the standard approach for setting exclusion limits in CMS analyses, allowing for a more conservative estimate and prevent false exclusion in the absence of sensitivity.

5.8. Statistical Modelling for the Analysis

This section presents the statistical techniques used in this analysis. with a detailed discussion on the the background modelling, signal shape parameterisation, and systematic uncertainty treatment. The previous section contained a short discussion of statistical techniques applied in LHC analyses, with a full treatment lying beyond the scope of this thesis. Therefore, the reader is now in a position to understand the specifics of this analysis.

While most of the statistical framework described in this section applies universally to all five methods that were part of CMS-EXO-22-026, certain aspects of the limit setting procedure differ between methods. In particular, the weakly supervised techniques require a distinctly different approach since these methods are sensitive to the presence (or absence) of signal. As these differences are extensively documented in the analysis publication and internal technical notes, and do not directly impact the VAE-QR method which forms the central focus of this thesis, they are not explored in detail here. The section concludes with a discussion on the strategy for interpreting potential excesses that might be observed following the unblinding of the data, in accordance with the established statistical framework at the LHC.

5.8.1. Modelling the QCD Background

The dijet invariant mass spectrum from QCD multijet processes is characterised by a smooth, monotonically decreasing function. This smoothly falling background shape has been studied in a previous CMS analysis [156]. In line with the data driven approach of this analysis, this parametrisation does not rely on a data control region or simulated samples for background modelling.

The background parameterisation employs the following functions, where the number of parameters can be either 2, 3 or 4:

$$\begin{aligned}
\frac{dN}{dm_{jj}} &= \frac{P_0(1-m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2}} \\
\frac{dN}{dm_{jj}} &= \frac{P_0(1-m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2+P_3\log(m_{jj}/\sqrt{s})}} \\
\frac{dN}{dm_{jj}} &= \frac{P_0(1-m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2+P_3\log(m_{jj}/\sqrt{s})+P_4\log(m_{jj}/\sqrt{s})^2}}
\end{aligned} \tag{5.41}$$

where m_{jj} represents the dijet invariant mass, $\sqrt{s}$ denotes the center-of-mass energy of the collision, P_0 serves as a normalisation parameter for the probability density function, and parameters P_i, $i = 1, 2, 3, 4$ determine the exact shape of the distribution. In the fitting procedure, these parameters are allowed to float freely during the optimisation process.

The necessary complexity of the background model is determined through Fisher's F-test [157]. This statistical procedure begins with the simplest viable model (the 2-parameter function) and evaluates whether additional complexity is justified by improved goodness-of-fit. While models with more parameters will invariably fit the data at least as well as simpler models, the F-test quantifies whether this improvement is statistically significant. The test statistic is defined as:

$$\text{Fisher}_{12} = \frac{\frac{\chi_1^2-\chi_2^2}{p_2-p_1}}{\frac{\chi_2^2}{n-p_2}} \tag{5.42}$$

where χ_i^2 represents the χ^2 of model i, p_i denotes the number of parameters in model i ($p_2 > p_1$), and n corresponds to the number of data points. A low confidence level (CL $< 10\%$) for this statistic indicates that the more complex model provides a significantly better description of the data, justifying the inclusion of additional parameters.

It is important to note that increasing model complexity may occasionally introduce parameter correlations that unreasonably inflate the uncertainty in the fitted background shape. To mitigate this issue, fits are only considered valid if they yield a relative uncertainty of less than 50% when evaluated at the resonance mass under investigation. Fits failing to meet this criterion are excluded from consideration in the F-test. In such cases, the F-test compares only those parameterisations whose fits satisfy this uncertainty threshold.

For the purpose of visualizing the fits and computing χ^2 values that avoid bins with zero entries, a specific binning scheme is employed. This dijet binning approximates the detector resolution as documented in a previous study [156], with slightly larger bin sizes at higher masses to account for decreasing energy resolution. The dijet binning scheme (in units of GeV) is given below:

$$\begin{aligned}\text{BINS} = [&1460, 1530, 1607, 1687, 1770, 1856, 1945, 2037, 2132, 2231, 2332,\\ &2438, 2546, 2659, 2775, 2895, 3019, 3147, 3279, 3416, 3558,\\ &3704, 3854, 4010, 4171, 4337, 4509, 4700, 4900, 5100, 5300,\\ &5500, 5800, 6100, 6400, 6800]\end{aligned} \tag{5.43}$$

For statistical evaluation, if the fit range is constrained compared to the full range (for instance, if the selected data extends only up to 5000 GeV), dijet bins outside this range are excluded from consideration. Additionally, to ensure statistical robustness in the χ^2 computation, bins containing fewer than 5 events are merged with adjacent bins until all bins contain at least 5 events.

The VAE-QR method performs the search for resonant anomaly signatures specifically within the mass range of $[1800, 6000]$ GeV, in intervals of 100 GeV using signal shapes that are described in detail in the following section.

5.8.2. Signal Shape modelling

An essential component of the bump-hunt search methodology is the creation of a template that adequately represents the expected signal shape. For resonance searches in the dijet mass spectrum, this shape is derived from fits to MC signal samples using a Double Crystal Ball (DCB) function [158]. The DCB function combines a Gaussian core, which accounts for detector resolution effects, with power-law tails on both sides to capture parton distribution function effects. This composite distribution is particularly suited for modelling resonance peaks in hadron collider experiments where asymmetric tails are common.

The complete functional form of the DCB function is as follows:

$$f(x) = N \begin{cases} A_1 \times (B_1 - \frac{x-\mu}{\sigma})^{-n_1}, & \frac{x-\mu}{\sigma} \leq -\alpha_1 \\ \exp(-\frac{(x-\mu)^2}{2\sigma^2}), & -\alpha_1 < \frac{x-\mu}{\sigma} < \alpha_2 \\ A_2 \times (B_2 + \frac{x-\mu}{\sigma})^{-n_2}, & \frac{x-\mu}{\sigma} \geq \alpha_2 \end{cases} \tag{5.44}$$

with

$$A_i = \left(\frac{n_i}{|\alpha_i|}\right)^{n_i} \exp\left(-\frac{\alpha_i^2}{2}\right), \quad B_i = \frac{n_i}{|\alpha_i|} - |\alpha_i| \tag{5.45}$$

where x represents the variable of interest m_{jj}, N is a normalisation constant, and μ, σ, α_1, α_2, n_1, and n_2 are parameters characterising the signal shape. The parameters μ and σ define the central peak position and width of the Gaussian core, respectively. The transition points between the Gaussian core and the power-law tails are controlled by the α parameters, with α_1 defining the transition point on the left side and α_2 the transition point on the right side of the distribution. The power-law exponents n_1 and n_2 determine how rapidly the tails fall off, with larger values of n_i resulting in faster-falling tails, while smaller values produce more extended tails. These parameters are particularly important for modelling the asymmetric

nature of resonance peaks, where initial-state radiation and parton distribution functions can significantly affect the low-mass tail.

When performing the profile likelihood fit to the data distribution, the power-law parameters (α_1, α_2, n_1, and n_2) are held fixed to values determined from the interpolated signal shapes, while the normalisation parameter N remains freely floating, as it directly corresponds to the expected signal cross section.

For the search in actual data, a generic signal shape is employed to maintain sensitivity across a wide range of mass hypotheses. This generic shape is derived from the $X \rightarrow YY'$ model with M_Y = 80 GeV and $M_{Y'}$ = 170 GeV as the reference signal. This particular model is selected because it produces daughter jets of reasonable mass that are light enough to result in fully merged decays (i.e., daughters fully contained within a single AK8 jet), leading to a characteristic signal shape with an extended low-mass tail.

The interpolation strategy for generating signal shapes at arbitrary mass points employs a combination of cubic splines and linear interpolation. This approach is similar to the methodology used in the diboson resonance search [159]. For the μ and σ parameters, which define the Gaussian core, cubic spline interpolation is used, whereas for the power-law parameters A_i and B_i, a simpler linear interpolation was observed to be sufficient.

This interpolation procedure generates signal templates at 100 GeV intervals spanning the range from 1800 to 5800 GeV. The parametric shape uncertainties that arise from uncertainties in the JES and JER are incorporated by applying variations to the peak (1.2%) and width (8%) of the Gaussian, respectively. These variations account for the experimental uncertainties in the jet reconstruction process that could affect the signal shape.

The effectiveness of this interpolation strategy is validated by removing the 3 TeV signal from the set of input samples used for parameter interpolation. The resulting interpolated signal shape at 3 TeV is then compared with the actual shape derived from direct fitting of the 3 TeV signal, as illustrated in Figure 5.17. Only minor differences, as quantified by the values of the signal shape parameters, are observed between the interpolated and actual signal shapes, confirming its robustness. The complete set of signal shapes generated across the full mass range through this interpolation procedure is presented in Figure 5.18.

With the background (Equation 5.41) and signal (Equation 5.44) being defined in this section, one is now in a position to begin constructing the likelihood of Equation 5.21.

5.8.3. Handling of Systematic Uncertainties

Systematics uncertainties were previously described in Section 5.7.3, with this section focusing on the specifics required for this analysis. A distinctive feature of this analysis is its *data-driven* nature, which is different from the conventional approach encountered in most CMS analyses. Here, the background estimation is derived directly from the data, eliminating the reliance on background simulation and its associated modelling uncertainties. The background normalisation and shape parameters that appear in Equation 5.41 are straightforward to obtain, using a uniform prior during the fit procedure (see Equation 5.27)

For the limit-setting procedure, where specific signal hypotheses are tested, additional systematic uncertainties come into play. These arise from the modelling of individual signal models, with particular attention required for their substructure characteristics. The accurate

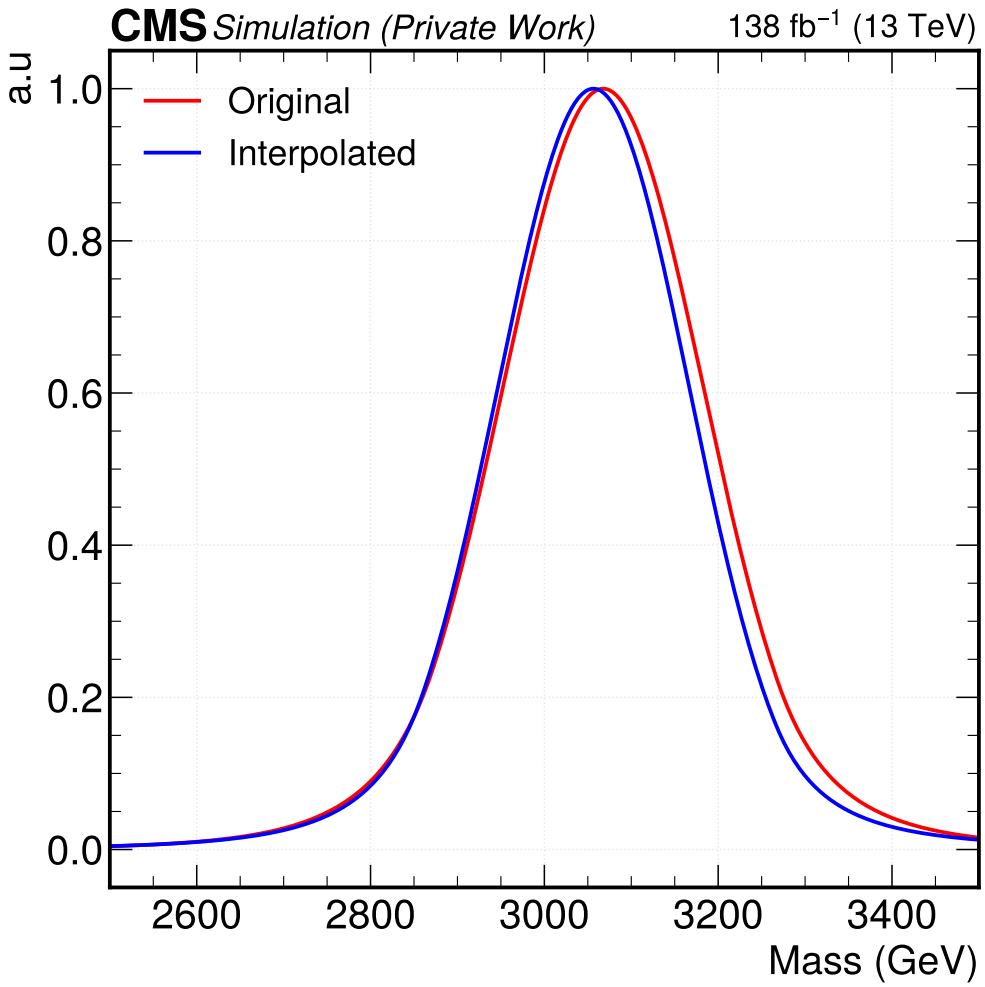


Figure 5.17.: Validation of the signal shape interpolation procedure. The interpolated signal shape at 3 TeV is compared to the actual shape derived from a direct fit to the 3 TeV signal resonance. The 3 TeV sample was excluded from the set used to generate the interpolation parameters.

description of jet substructure is crucial for the proper modelling of signal efficiency and acceptance, especially since no SM proxies exist for calibrating jets with more than 3 prongs. These effects necessitate dedicated studies to understand how detector resolution, calibration uncertainties, and reconstruction algorithms affect the signal modelling.

As is standard practice within the CMS Collaboration, the profile likelihood method (see Section 5.7.2) is used for hypothesis testing, whereas the CL_s procedure (see Section 5.7.8) is used for limit setting. All fits are implemented using the COMBINE [160] Tool.

In the forthcoming sections, the individual sources of systematic uncertainties affecting this analysis will be discussed in detail, focusing first on signal shape uncertainties followed by additional modelling uncertainties (the so-called signal normalisation uncertainties) relevant for the specific signal hypotheses on which limits are set.

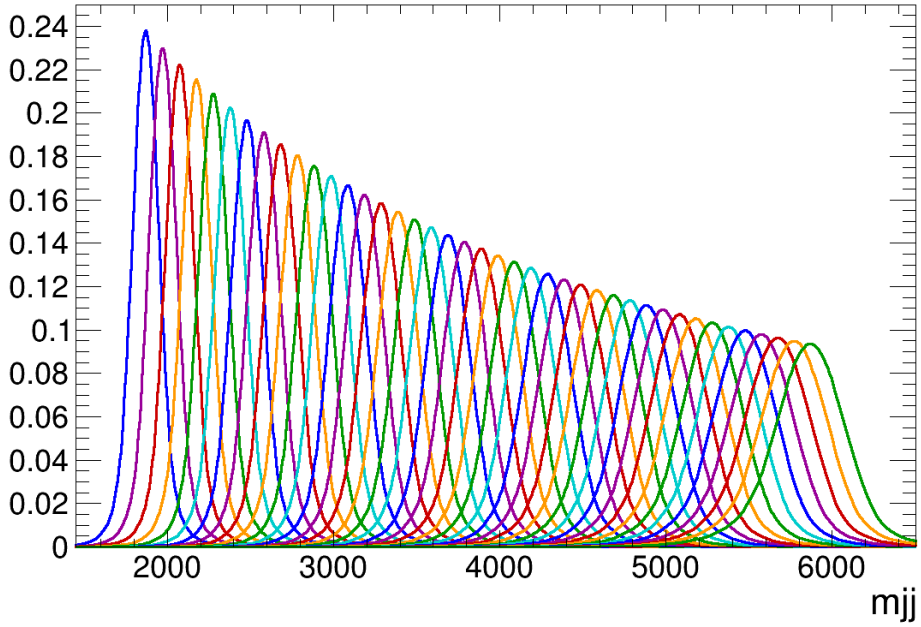


Figure 5.18.: The complete set of interpolated signal shapes used for the resonant anomaly signature search procedure, shown here over the entire mass range of 1800 to 5800 GeV in steps of 100 GeV. Each curve represents the expected signal shape for a specific resonance mass hypothesis. The progression shows how the peak position shifts and width increases with increasing resonance mass, reflecting the combined effects of detector resolution and parton distribution functions.

5.8.4. Event Weights for modelling Systematic Uncertainties

In practice, systematic uncertainties are modeled by associating a set of event weights with each simulated event, where each weight corresponds to a specific variation of a systematic parameter. This approach allows the impact of systematics to be evaluated by reweighting the same base events rather than generating new simulated samples for each variation, and significantly reduces computational requirements by avoiding repeated MC event generation and detector simulation for each such variation.

5.8.5. Signal Shape Uncertainties from JES and JER

The JES represents the collection of correction factors applied to jet energy measurements obtained from the CMS detector. These corrections account for various detector-specific effects, such as calorimeter effects, differences in detector responses, and calibration uncertainties. By applying these corrections, the measured jet energies are accurately related to their corresponding true energies. The JER, on the other hand, quantifies the intrinsic fluctuations and resolution limitations of the jet energy measurement. Specifically, JER characterises the width of the jet energy response distribution while JES defines the position of the jet energy peak.

To systematically quantify these uncertainties, a variation procedure is implemented wherein the four-momenta of the jets comprising the dijet system are varied according to the JES and JER uncertainties. For the JES variations, the jet energies are shifted upward and downward by one standard deviation of the known JES uncertainty, which typically ranges from 1% to 3% depending on the p_T and η. For the JER variations, the resolution smearing applied

to simulation is adjusted by $\pm 1\sigma$ of its uncertainty. This procedure yields an uncertainty of approximately 1% on the μ parameter and 3.5% on the σ parameter.

Following these variations, the signal mass distributions fits are repeated with the DCB function of Equation (5.44). The differences in the fitted μ and σ values between the nominal template and the up/down variations are recorded as the systematic variations on these parameters.

In the signal-plus-background fitting procedure, these uncertainties are incorporated by allowing the μ and σ parameters to float within Gaussian constraints centered on their nominal values, with widths corresponding to the derived uncertainties. This approach effectively propagates the JES and JER uncertainties through to the final result.

For the generic signal template used as a model-independent search tool, this uncertainty assessment is repeated at several reference mass points (2, 3, and 5 TeV), revealing comparable uncertainties with slightly larger values observed for higher mass signals due to the decreased detector resolution at extreme energies. These derived uncertainties are subsequently applied to all mass points through interpolation, ensuring a continuous treatment of systematic effects across the entire mass spectrum investigated in the analysis.

5.8.6. Signal Normalisation Uncertainties

The signal normalisation uncertainties primarily affect the limit-setting, since it affects the efficiency of the signal selection within each category using the VAE-QR. As mentioned in Section 5.8.3, the COMBINE tool is used for the mathematical implementation. These uncertainties are correlated among all years of data taking. A short discussion of the important sources of normalisation uncertainty, beginning with the jet substructure modelling (the so-called Lund Jet Plane) uncertainty, is provided now:

- **Jet Substructure Modelling Uncertainties:** It is not trivial to set limits on signal models with more than three prongs, due to the absence of a SM proxy. The SM only provides adequate proxies for one-, two-, and three-pronged jets (light quarks, boosted W/Z bosons, and top quark decays, respectively). To address this limitation, a novel approach [161, 162] that computes the ratio between data and MC in the Lund Jet Plane [163] was developed [2] to compute per-prong substructure correction weights, for the purpose of this analysis. In essence, this method derives a correction factor using a ratio between data and simulation in the aforementioned plane. The Lund Jet Plane is a two dimensional representation, with $\ln k_T$ on the y-axis and $\ln(1/\Delta)$ on the$x$-axis, that shows the splittings in the showering procedure of a given jet. Here, k_T and Δ represent the relative p_T and the angular separation of an emitted particle, with respect to its emitter. The procedure derives weights for each prong within a jet by reclustering it to find the individual subjets. This procedure can then be used to build up the corrections for multi-prong resonances with no SM proxies, by ensuring that each prong is contained within its own separate subjet. It is important to note that several factors contribute to the final uncertainty, as summarised below:
  - Systematic uncertainties arising during the Lund Jet Plane ratio calculation

[2] Primarily due to the excellent work of collaborator Oz Amram (Fermilab)

  - Statistical uncertainty arising from the measurement of the ratio in the Lund Jet Plane, which is modelled by randomly sampling a fixed number (100) of values from each bin in the 2D plane, where the bin is modelled as a standard Gaussian centred at the nominal bin value.
  - Since quarks that are close to the jet boundary of an AK8 jet may produce hadrons that lie outside, a variation of ±1 to the number of prongs found by the reclustering algorithm is carried out, and the results included as variations to the event weight
  - Additionally, when the (re-)clustering algorithm fails to match a given subjet within the AK8 jet to a generator-level quark, a conservative approach is chosen, where the event weight is varied by the maximum possible factor of 5.
  - The Lund Jet Plane ratio is agnostic to quark flavour, but this assumption is not always true, since colour reconnection effects can distort it depending on whether the initiating quark originated from a given signal or a known *standard candle* signal. Therefore for each signal model, the ratio of the Lund Plane for the signal is compared to that of the two-pronged W boson decay, with this *new* ratio being used as variations to the event weights. This is the **dominant** source of uncertainty (see Figure 5.19b).

- **Jet energy/mass scale and resolution:** Uncertainties addressed by varying the four-momentum vectors of the AK8 jet within uncertainty limits (5% for jet mass scale, 8% for resolution) following the official recommendations [164] from the CMS Collaboration.
- **L1 trigger prefiring:** Accounts for ECAL cluster time shift in 2016-2017 data predominantly at high p_T and η [165]. Event weights are calculated by varying the probability of AK4 jets or photons having caused this *prefiring*, up and down by 20% [166].
- **Luminosity:** Total integrated luminosity of Run 2 is known to 1.6% precision, this being applied as a normalisation uncertainty [167, 168, 169]. Its effect was found to be negligible.
- **Renormalisation and Factorisation scale:** Cross-section uncertainties estimated by varying these scales up and down by factor of two, both separately and together, with anti-correlations being physically excluded.
- **Parton shower modelling:** Uncertainties in parton shower simulation derived by varying the renormalisation scale by a factor of two, up and down. For the purpose of this analysis, only the initial-state radiation factor counts.
- **Pileup reweighting:** Simulation adjusted for pileup interactions via event weights derived from inelastic collision cross-section, with the uncertainty being calculated by shifting the minimum bias cross-section up and down by 4.6%, and thereafter being propagated to the event weights.
- **Parton distribution functions:** PDF uncertainties incorporated through event weights based on fit uncertainties, with the standard deviation used to vary the nominal weight and produce a reduced set of weights.
- **Top quark p_T reweighting:** Corrections applied to account for mismodeled top quark p_T spectrum due to missing higher order effects, with weights derived from differences between unfolded data and MC generator predictions [170].

5.8.7. Incorporating systematic uncertainties into VAE-QR

The integration of systematic uncertainties into the VAE-QR methodology warrants careful consideration before proceeding to the fitting strategy. As previously established, VAE-QR employs a QR to define multiple signal-enriched categories. When implementing the final statistical fit in these categories, with the rationale for these categories being elaborated upon in the forthcoming Section 5.8.8, systematic uncertainties must be rigorously accounted for. In accordance with established practice within the CMS Collaboration, the COMBINE tool is employed to perform hypothesis tests and compute exclusion limits. The inherently data-driven nature of VAE-QR presents a major advantage, as variations in signal shape or normalisation arising from systematic uncertainties do not affect the VAE or the QR, both of which are trained exclusively on data. As a result, re-training is **not** necessary. Nevertheless, signal normalisation uncertainties require meticulous treatment, as they potentially alter the event distribution across categories. Consequently, for each systematic uncertainty described in Section 5.8.6, the selection efficiency under both upward and downward variations is calculated and incorporated as an input to the COMBINE tool, maintaining consistency with the log-normal treatment of normalisation uncertainties detailed in Section 5.7.6 (and formalised in Equation 5.29). The quantitative impact of these variations is illustrated in Figure 5.19 for the Lund Jet Plane systematic uncertainties across both the single-category Q_{90} and the three categories utilised in the statistical fit (refer to Section 5.8.8). Additionally, Figure 5.20 depicts the constituent contributions to the Lund Jet Plane uncertainties for the single category Q_{90}. These plots are shown for the specific example of the signal model $X \to YY'$ with parameters $M_X = 3$ TeV and $M_Y = M_{Y'} = 170$ GeV.

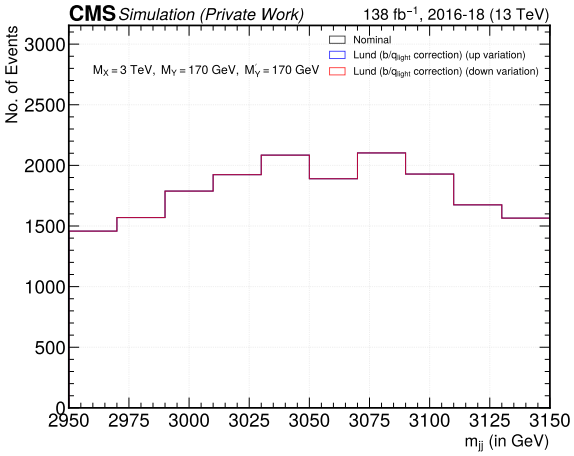
5.8.8. Choice of categories and the final fit strategy for VAE-QR

The VAE-QR methodology divides the SR into multiple disjoint categories using the QR. Most importantly, in the absence of signal, all categories have a smoothly falling m_{jj} distribution described by Equation 5.41. Consequently, the differences in autoencoder reconstruction loss between signal and QCD background result in these categories having varying signal-to-background (S/B) ratios should any such signal be present.

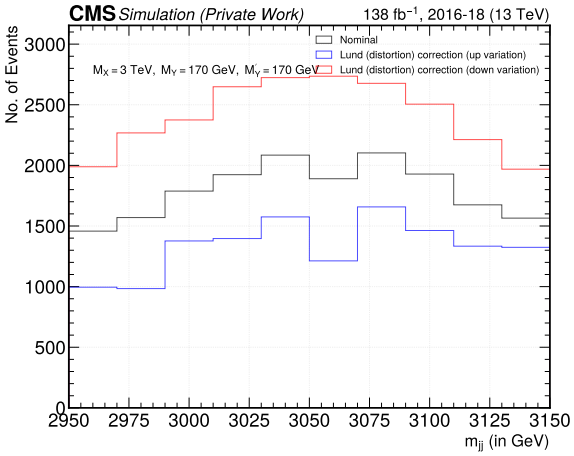
While the search procedure employs solely the 10% most anomalous events (denoted as the Q_{90} category) to maximise discovery potential, the limit-setting framework leverages a more nuanced multicategory approach to enhance statistical sensitivity. This framework utilises three distinct categories defined by the QR percentile boundaries:

- Category 1: Q_{99}, comprising the most anomalous 1% of events, representing the highest signal purity.
- Category 2: Q_{95-99}, which contains events with anomaly scores between the 95^{th} and 99^{th} percentile.
- Category 3: Q_{90-95}, encompassing events with anomaly scores between the 90^{th} and 95^{th} percentile, offering statistical power marginally higher than Category 2.

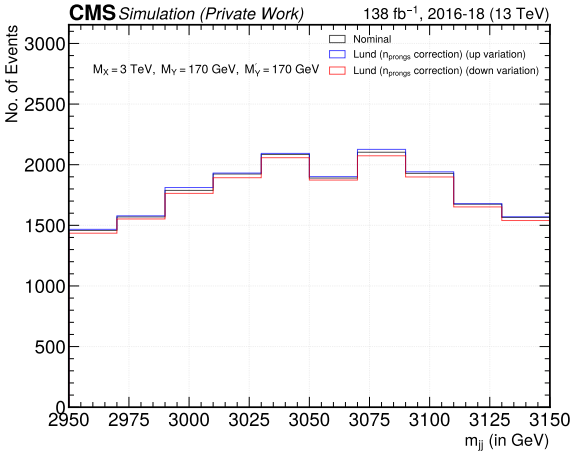
Category 3 effectively constrains the background parameterisation for all categories, capitalizing on the shape invariance across categories inherent in the QR methodology. The requisite number of background function parameters in Equation 5.41 is determined by fitting the events



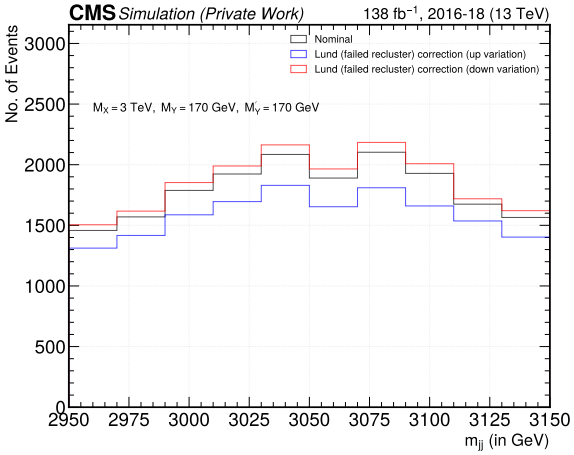
(a) Variations due to the different shower pattern of a b-quark



(b) Variation due to distortion of the Lund Plane



(c) Contribution from the prong variation



(d) Variation due to failed re-clustering

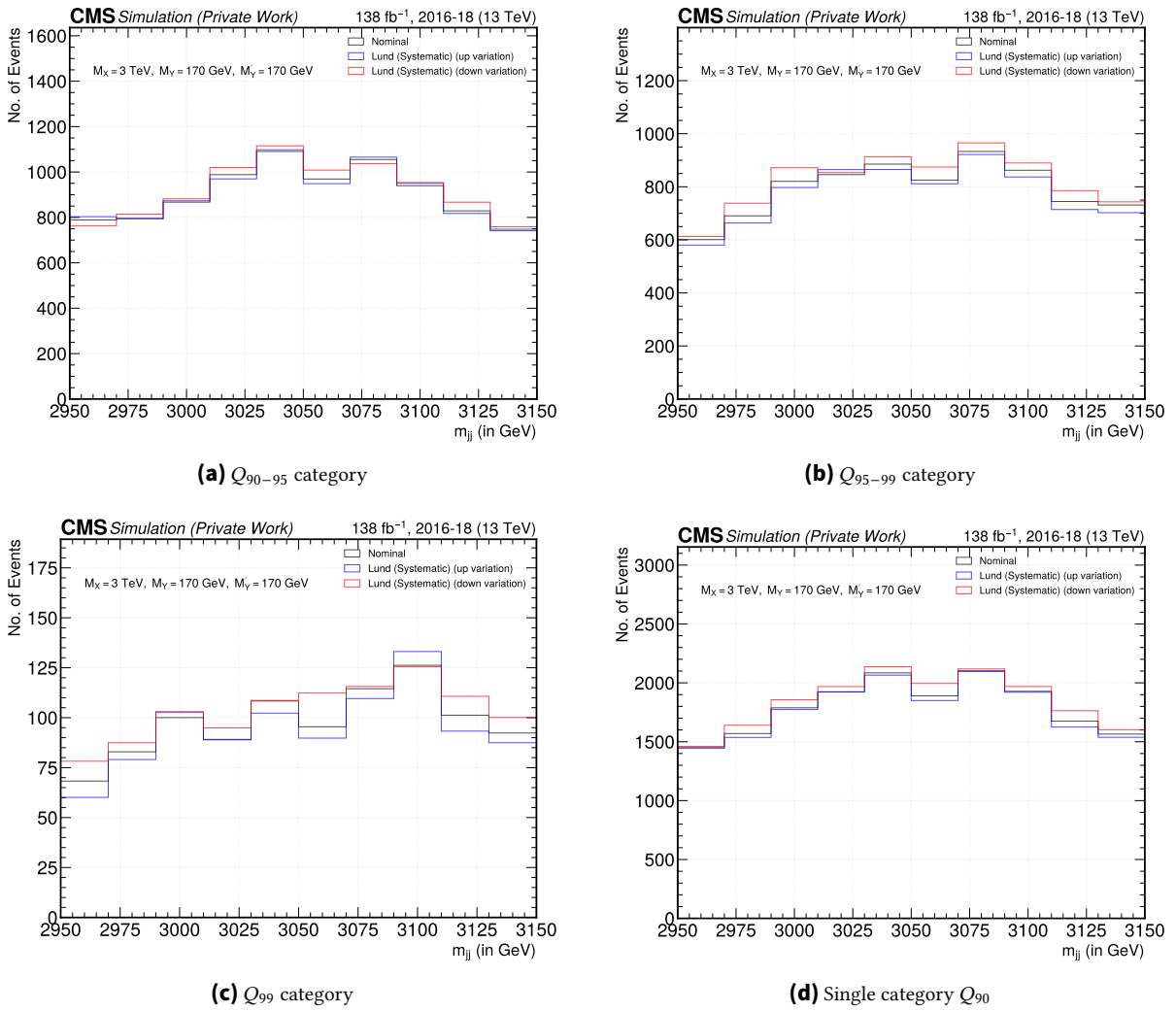
Figure 5.19.: Contribution of Lund Jet Plane systematic uncertainty components on the selection efficiency of the single category Q_{90}, which contains the top 10% anomalous events for the $X \rightarrow YY'$ signal model with $M_X = 3$ TeV and $M_Y = M_{Y'} = 170$ GeV.

in Category 3 and subjecting the results to a Fisher F-Test to assess the statistical significance of additional parameters, as illustrated in Figure 5.21. The derived parameterisation is subsequently applied across all categories.

This is followed by a combined likelihood fit performed simultaneously across all three categories. The background function parameters are permitted to float freely within their uncertainties, while the background normalisation for Categories 1 and 2 are constrained to be predetermined fractions of Category 3's background normalisation (see Figure 5.16), which itself floats freely in the fit.

To validate the robustness of this statistical framework, three distinct fitting approaches were applied to a mock MC dataset equivalent to an integrated luminosity of 26.8 fb^{-1}without signal injection:

- The proposed multicategory simultaneous fit
- A single-category fit restricted to events exceeding the Q_{90} threshold
- An inclusive fit without the VAE-QR selections



(a) Q_{90-95} category

(b) Q_{95-99} category

(c) Q_{99} category

(d) Single category Q_{90}

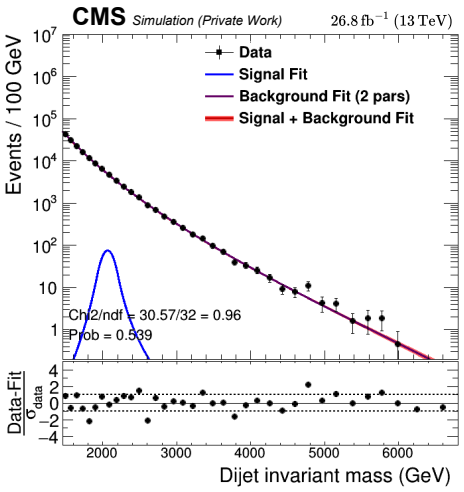
Figure 5.20.: Both upward and downward variations induced by the Lund Jet Plane systematic uncertainties across different categories for the $X \rightarrow YY'$ signal model with $M_X = 3$ TeV and $M_Y = M_{Y'} = 170$ GeV

All three approaches consistently indicated that a two-parameter formulation of the background function provides an adequate description of the background distribution, as evidenced by Figures 5.21 and 5.22. This validates the assumption that Category 3 can effectively constrain the background parameterisation for Categories 1 and 2 and ultimately enhance sensitivity. The signal injection studies described in Appendix B.2 provide further justification for the multi-category fit described herein, as they are shown to improve sensitivity for less boosted decay products of the parent resonance.

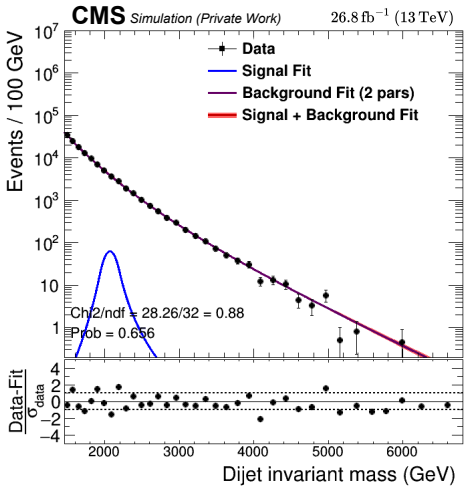
5.8.9. Final Hypothesis Tests and Limit Extraction

Having constrained the background using the high-statistic Category 3, and given the systematic uncertainties described in the previous two sections, all ingredients for performing a resonant anomaly hunt and setting limits on the chosen signal models are in place.

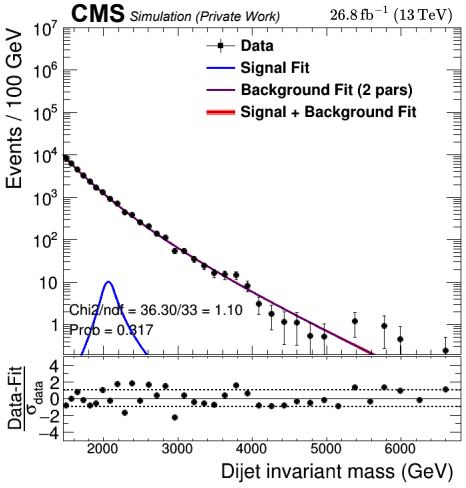
In statistical terms, the systematic uncertainties are incorporated into the analysis framework as nuisance parameters in the profile likelihood fit. The profile likelihood approach allows for the simultaneous fitting of the parameters of interest (signal strength) and the nuisance



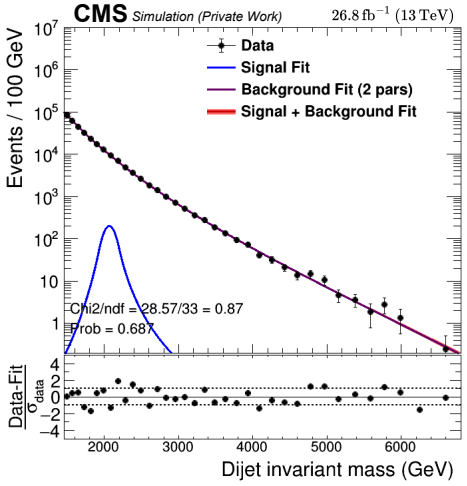
(a) Category 3 (Q_{90-95}): Contains the highest statistics and is used to constrain the background shape.



(b) Category 2 (Q_{95-99}): Background shape matches that of Category 3, differing only in normalisation.



(c) Category 1 (Q_{99}): Highest signal purity region with the same background functional form.



(d) Single-category approach with Q_{90} selection threshold, used for the search.

Figure 5.21.: Maximum likelihood fit for the different VAE-QR selection categories, on a mock MC dataset containing only QCD multijet events, corresponding to an integrated luminosity of 26.8fb^{-1}.

parameters representing the systematic uncertainties [148]. This technique effectively "profiles out" the nuisance parameters by finding their best-fit values conditional on each value of the parameter of interest.

As detailed in Section 5.7.2, the profile likelihood fit is used for the resonant anomaly search over the m_{jj} spectrum, while the CL$_s$ method at CL = 95% is used for limit setting. The output from this tool is an upper limit on the expected and observed signal strength parameter μ, with $\pm 1\sigma$ bands on the expected limit, for a total of four values.

The signal strength upper limit can be converted to an upper limit on the number of signal events by multiplying by the signal normalisation. This is a hyperparameter whose value can

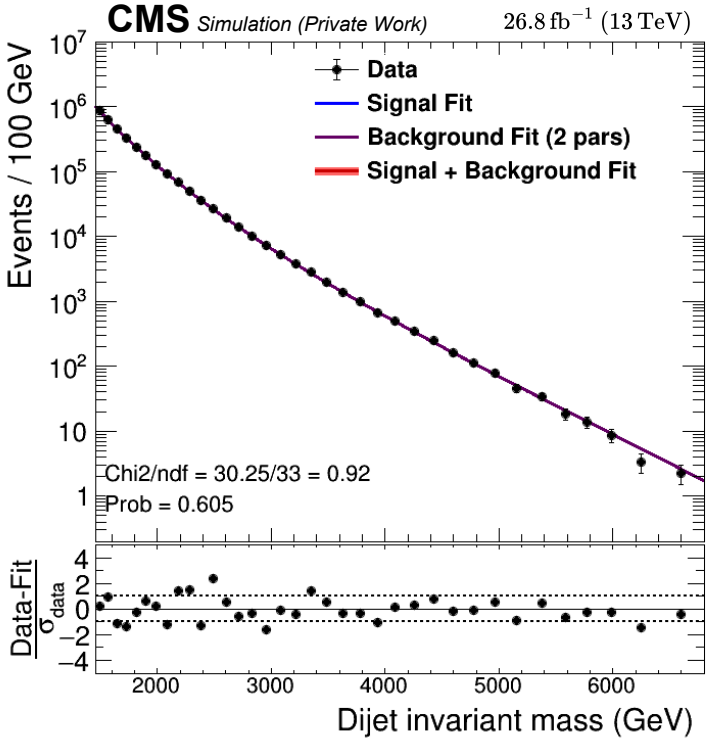


Figure 5.22.: Maximum likelihood fit to the inclusive MC mock dataset containing only QCD multijet events, without any QR selection applied.

be tuned to obtain stable fits. To convert this to a limit on the signal cross section, one must use the following formula:

$$\sigma_{\text{upper}} = \frac{N_{\text{upper}}}{\mathcal{L}_{int} \cdot \varepsilon_{\text{presel}} \cdot \varepsilon_{\text{sel}}} \tag{5.46}$$

The first two terms of Equation 5.46 are already known, these being the integrated luminosity $\mathcal{L}_{\text{int}} = 138\text{fb}^{-1}$ and the efficiency of the preselection criteria $\varepsilon_{\text{presel}}$ defined in Section 5.2.5 and Table 5.5. However, for the weakly supervised methods that make up this analysis, the calculation of the third term, which is the selection efficiency ε_{sel} is not trivial, since this number is dependent on the presence (or absence) of signal. On the other hand, for VAE-QR, the QR approach to selection of categories makes this calculation trivial, allowing one to directly convert the output from the Combine tool to the exclusion limits on the observed and expected signal cross sections, along with the associated uncertainty bands on the expected limits. Additionally, it is shown in Appendix B.2 that the QR itself is robust against signal contamination to a moderate degree. The VAE itself is trained on individual jets (that is to say, 2 jets per event) and does not possess an event-level view of the data in the first place. In addition, it is trained on jets from a specially constructed data control region (see Section 5.4) which is expected to have minimal signal contamination. Therefore, both individual components of VAE-QR are stable against the presence of small amounts of signal, which makes the final procedure for computing limits much simpler compared to the other methods present in the CMS analysis.

To conclude, before proceeding to the method validation and presentation of results, this discussion of the statistical framework was necessary to establish the statistical foundation

of the analysis. Two statistical approaches were implemented in this study, each serving a different purpose within the broader search strategy.

Firstly, a profile likelihood fit utilising highly generic interpolated signal shapes was employed to perform a model-independent search for resonant anomalies across the m_{jj} spectrum. This search was systematically conducted in the range of 1800 to 5000 GeV at 100 GeV intervals, allowing for comprehensive coverage. The technical implementation was done using the RooFit framework [153]. This software uses the minimisation capabilities of MINUIT2 and MiGrad to ensure reliable convergence [171].

Secondly, a dedicated limit-setting procedure was established to constrain specific signal models in the absence of significant excesses. This approach incorporated three distinct selection categories derived from the VAE-QR methodology, thereby optimising the sensitivity to signals with varying jet substructure characteristics. The statistical interpretation follows the CL_s technique, which has been adopted as the standard within the CMS Collaboration due to its robustness against statistical fluctuations in regions of low sensitivity [154]. The practical implementation of this procedure was performed using the CMS Combine tool [160].

With this statistical framework established, the subsequent sections will focus on the validation of the methodology and the presentation of the physics results in Section 5.9.

5.9. Results from the VAE-QR Methodology in this analysis

This section presents the results obtained with the VAE-QR method within the broader context of this analysis [51]. The discussion is structured to first address the significance scan findings, followed by a presentation of the limit-setting capability of the method across various signal models.

5.9.1. Significance Scan Results

A rigorous search was conducted across the m_{jj} range of [1800, 5000] GeV. **None** of the five model-agnostic methods employed in this analysis, including the VAE-QR approach, revealed **statistically significant excesses** above the SM background expectation. The most substantial deviation observed using the VAE-QR methodology was a 2.3σ local significance at a mass of 4.9 TeV, which remains well within the expected statistical fluctuations of the background-only hypothesis after accounting for the look-elsewhere effect [172].

The comprehensive significance scan performed using the VAE-QR method is illustrated in Figure 5.23. For comparative purposes, Figure 5.24 presents the corresponding significance scan from the inclusive analysis without any selection applied. The signal+background fit corresponding to the mass point exhibiting the highest significance (4.9 TeV) is shown in Figure 5.25 and the final background-only fit is shown in Figure 5.26.

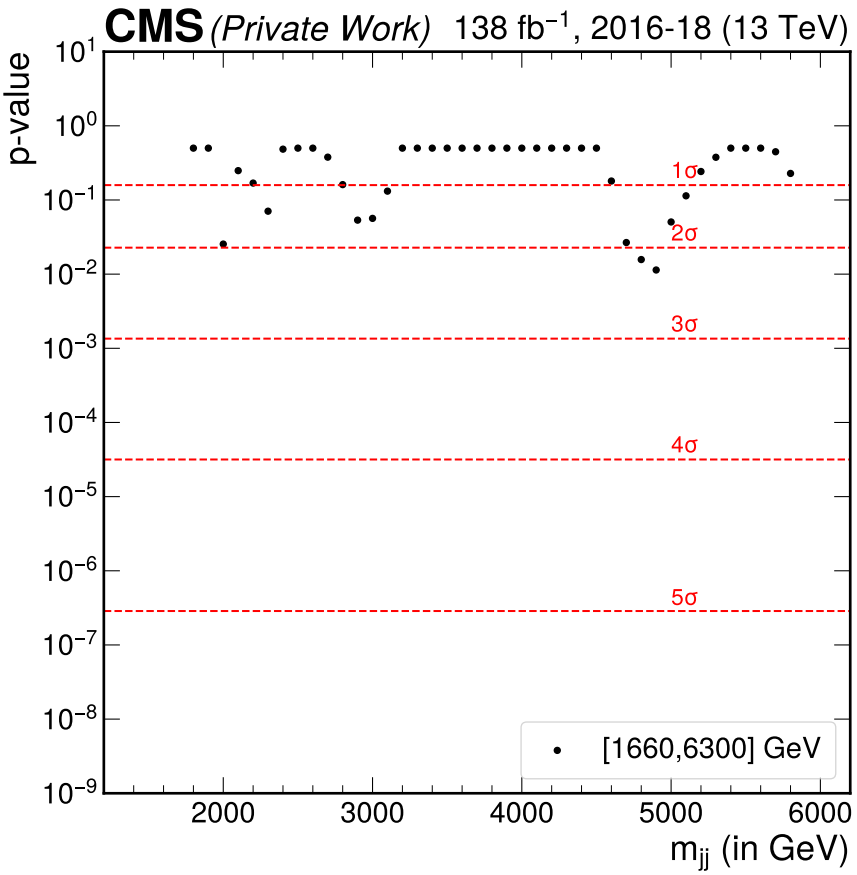


Figure 5.23.: Local significance (expressed in standard deviations) as a function of m_{jj} for the VAE-QR method. The horizontal dashed lines indicate the increasing significance levels. The highest local significance of 2.3σ occurs at approximately 4.9 TeV

5.9.2. Limit Setting on Various Signal Models

A distinguishing feature of this analysis lies in its model-agnostic approach, enabling the investigation of a diverse spectrum of theoretical scenarios without requiring explicit optimisation for specific signal topologies. The VAE-QR method is thereafter used to establish upper limits on the production cross section for an extensive array of signal models, a full list of which is given in Table 5.1.

It is particularly noteworthy that for the majority of these signal hypotheses, this analysis represents the first instance of experimental constraints being established, with the exception of the $W_{KK} \rightarrow WWW$ signal model, which has been previously investigated in a dedicated search [134]. This highlights the pioneering nature of this model-agnostic approach in exploring previously uncharted parameter spaces of BSM physics.

The 95% CL upper limits derived for signal masses of 2,3 and 5 TeV are presented in Figures 5.27, 5.28, and 5.29, respectively. Across the majority of signal hypotheses, the VAE-QR method demonstrates substantial improvement in sensitivity compared to the inclusive analysis, underscoring the efficacy of this anomaly detection technique. Nevertheless, it is important to acknowledge that the limits established through this model-agnostic approach are gener-

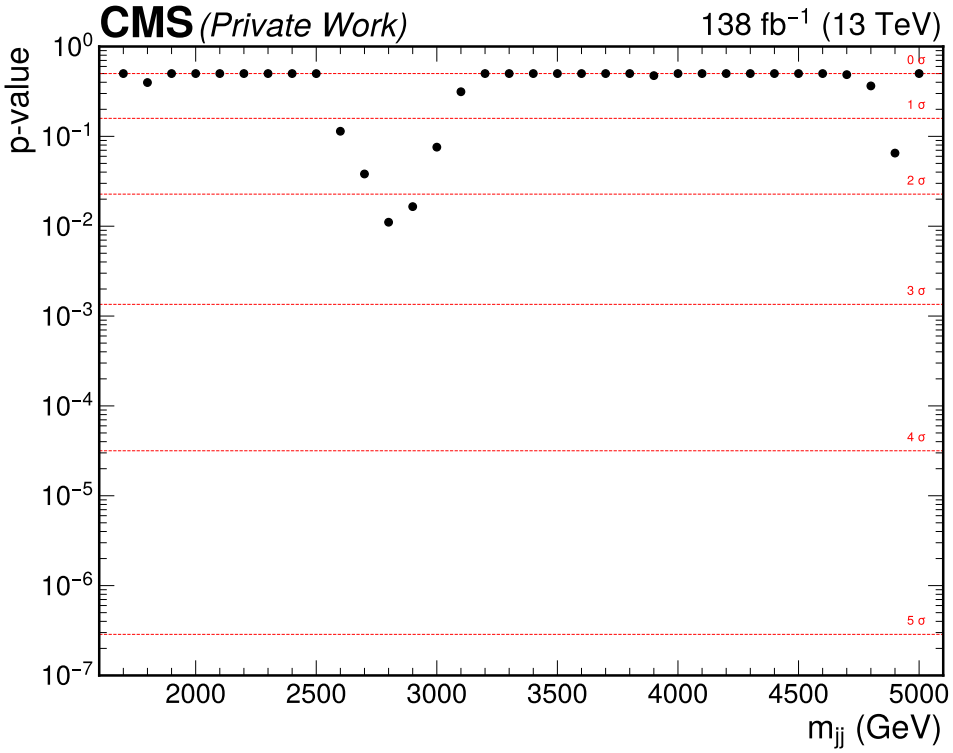


Figure 5.24.: Local significance (expressed in standard deviations) as a function of m_{jj} for the inclusive analysis without anomaly-detection selection

ally less stringent than those that would be attained through analyses specifically optimised for individual signal models, which represents an inherent trade-off between generality and specialised sensitivity.

5.9.3. Comparative Performance of Model-Agnostic Methods

To facilitate a comprehensive evaluation of the relative performance of different anomaly detection techniques, a set of benchmark signal models are selected, and the limit-setting capability of all five methods are systematically compared. The results of this comparative analysis are illustrated in Figures 5.30 and 5.31, It is observed that the presented anomaly detection methods consistently outperform the inclusive analysis across the investigated signal scenarios.

Table 5.6 enumerates the optimal anomaly detection method for each investigated signal topology, providing a quantitative foundation for examining the complementarity of different approaches. The VAE-QR method exhibits exceptional performance in scenarios where the parent resonance produces decay products of higher mass. This sensitivity to the mass scale of the overall event can be attributed to the foundational architecture of the VAE-QR, which is trained directly on basic kinematic three-vectors, which are the most elementary particle-level features available within the analysis, instead of the high level features used by the other anomaly detection methods.

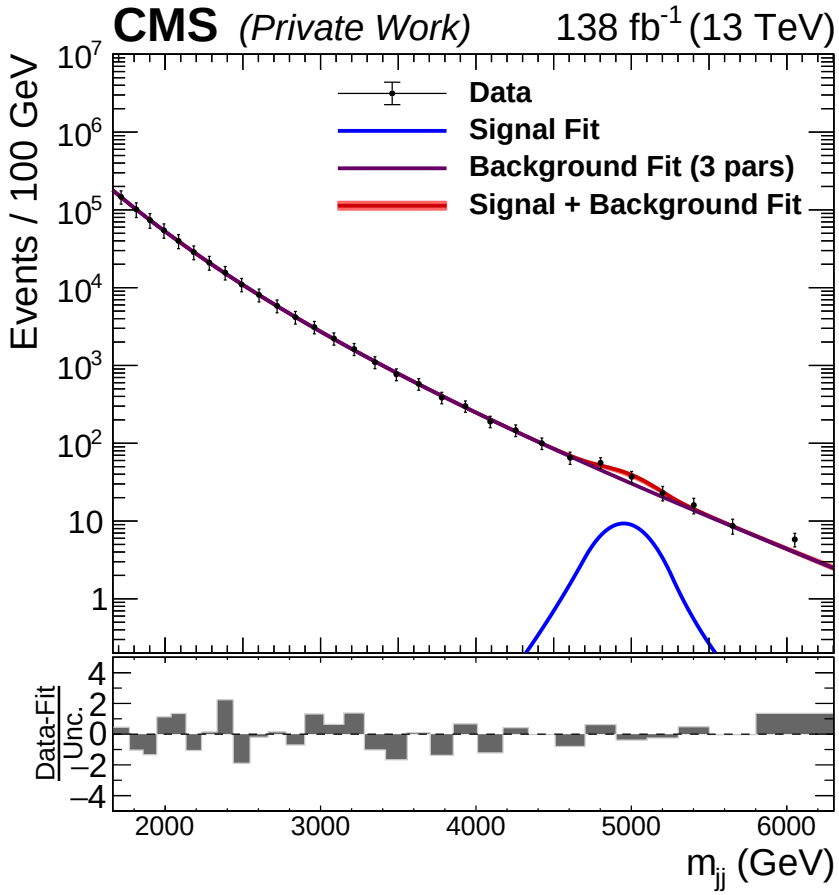


Figure 5.25.: Signal+Background fit, using the generic signal shapes described in Section 5.8.2 to the m_{jj} spectrum at the mass point of 4.9 TeV, which exhibited the highest local significance of 2.3σ in the VAE-QR analysis. The lower panel shows the pull distribution (data minus fit, divided by the statistical uncertainty)

5.10. Conclusion: Why VAE-QR?

The VAE-QR methodology occupies a distinct position within the ensemble of anomaly detection techniques employed in this analysis, primarily due to its highly generalised set of input features, and its unsupervised learning paradigm. Unlike other methods that incorporate higher-level features or require either a weakly or semi-supervised paradigm, the VAE-QR operates exclusively on the most fundamental representations of particle kinematics, specifically the three-momenta of the constituent particles of an AK8 jet. This approach avoids higher level features and complex preprocessing, instead allowing the neural network to extract relevant patterns directly from the raw data.

The **CMS-EXO-22-026** [51, 112] analysis represents a pioneering effort from the CMS Collaboration to develop and compare multiple model-agnostic methods within a unified framework and is a significant milestone in the search for BSM physics. The competitive performance demonstrated by the VAE-QR, despite its minimal assumptions and highly generalised nature, makes a compelling case for the usage of unsupervised anomaly detection techniques in HEP.

The particularly robust performance of the VAE-QR in scenarios involving high-mass decay products proves its inherent sensitivity to the characteristic energy scales of physical processes. This makes it valuable for probing the high-energy frontier, where new heavy resonances

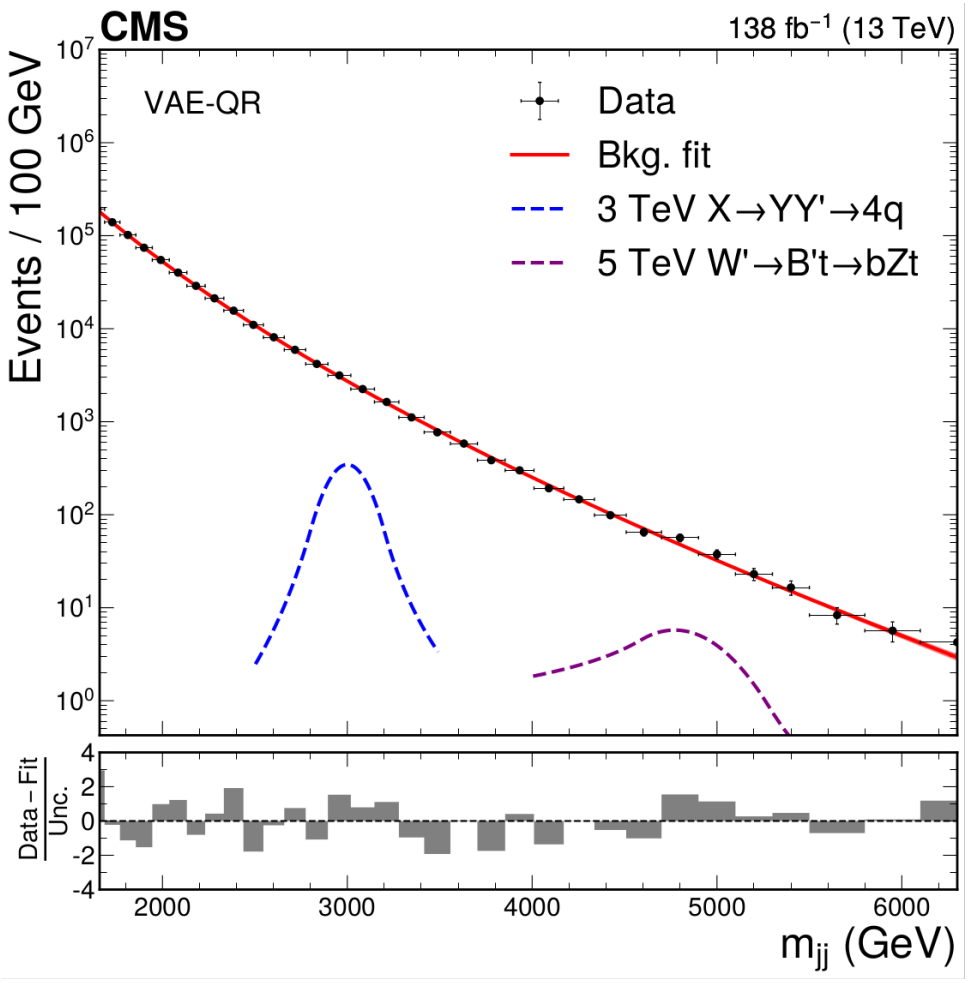


Figure 5.26.: Final background-only fit to the events of Category 3, as selected using the VAE-QR methodology. Two representative signal models of mass 3 TeV and 5 TeV are shown on the panel.

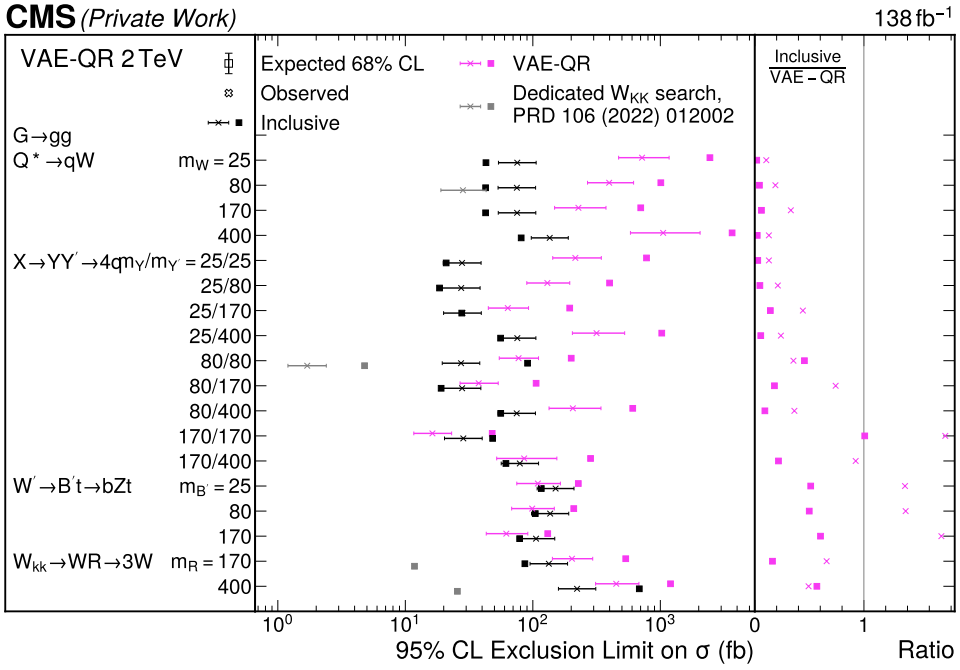


Figure 5.27.: Upper limits at 95% confidence level on the production cross section for various signal models with a resonance mass of 2 TeV, comparing the VAE-QR method with the inclusive analysis.

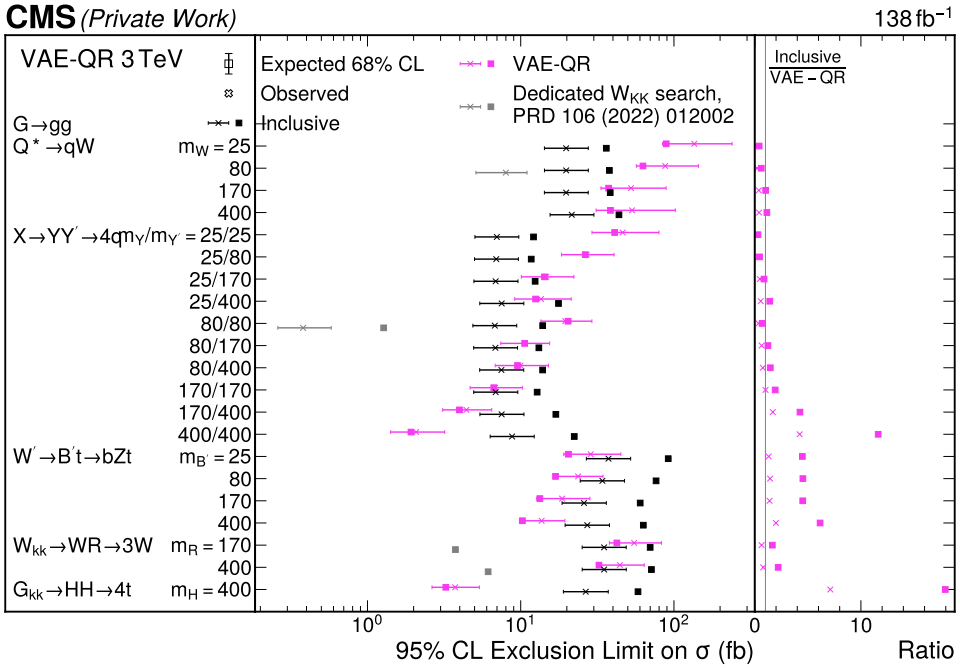


Figure 5.28.: Upper limits at 95% confidence level on the production cross section for various signal models with a resonance mass of 3 TeV, comparing the VAE-QR method with the inclusive analysis.

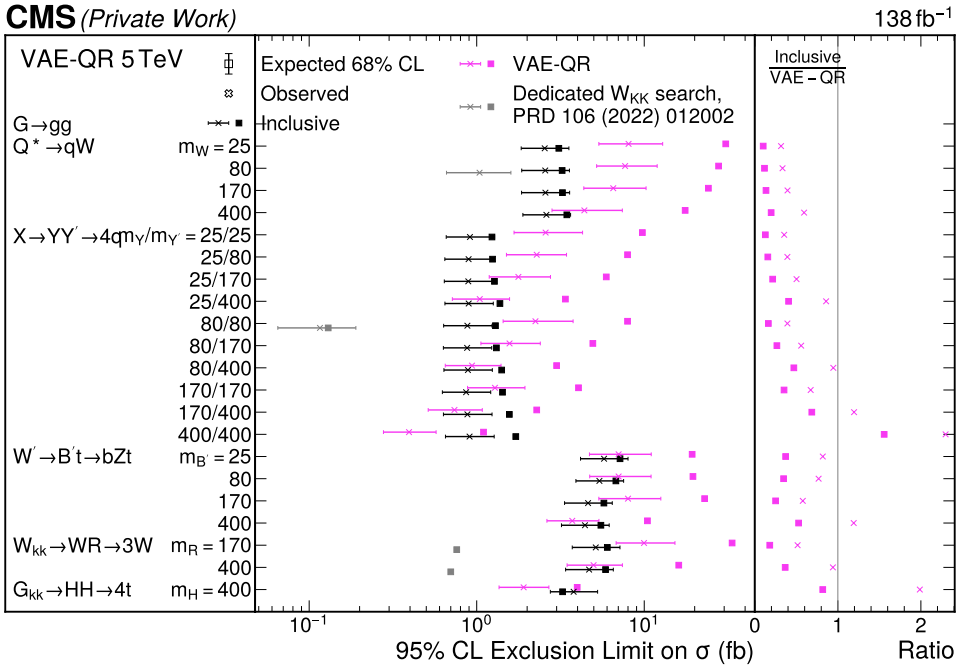


Figure 5.29.: Upper limits at 95% confidence level on the production cross section for various signal models with a resonance mass of 5 TeV, comparing the VAE-QR method with the inclusive analysis.

might theoretically be present. Furthermore, the model-agnostic nature of this approach allows one to target unconventional signal topologies that might be overlooked by more focused searches, thereby expanding the discovery potential of the CMS Experiment.

The reader is encouraged to look at Appendix B.3 for a short description of the other methods that formed part of the CMS-EXO-22-026 analysis and the complementarity of these approaches, which is not included within the scope of the main thesis itself.

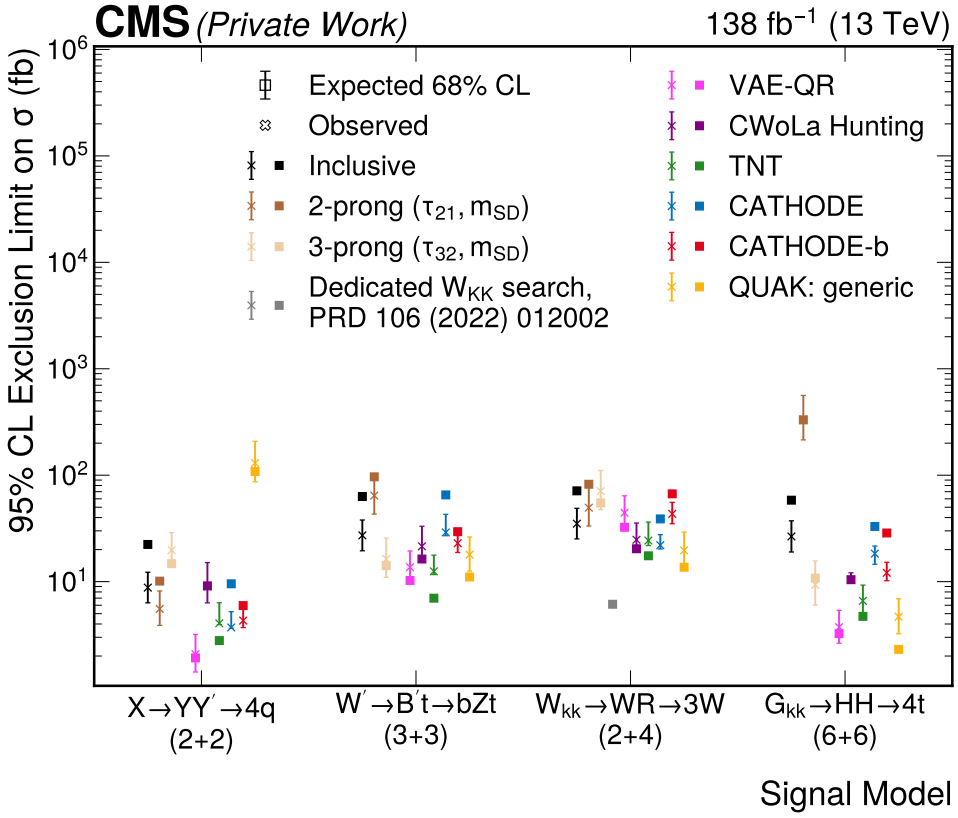


Figure 5.30.: Comparison of upper limits on the production cross section for benchmark signal models at 3 TeV, showing the relative performance of all five model-agnostic methods implemented in this analysis, alongside the inclusive approach.

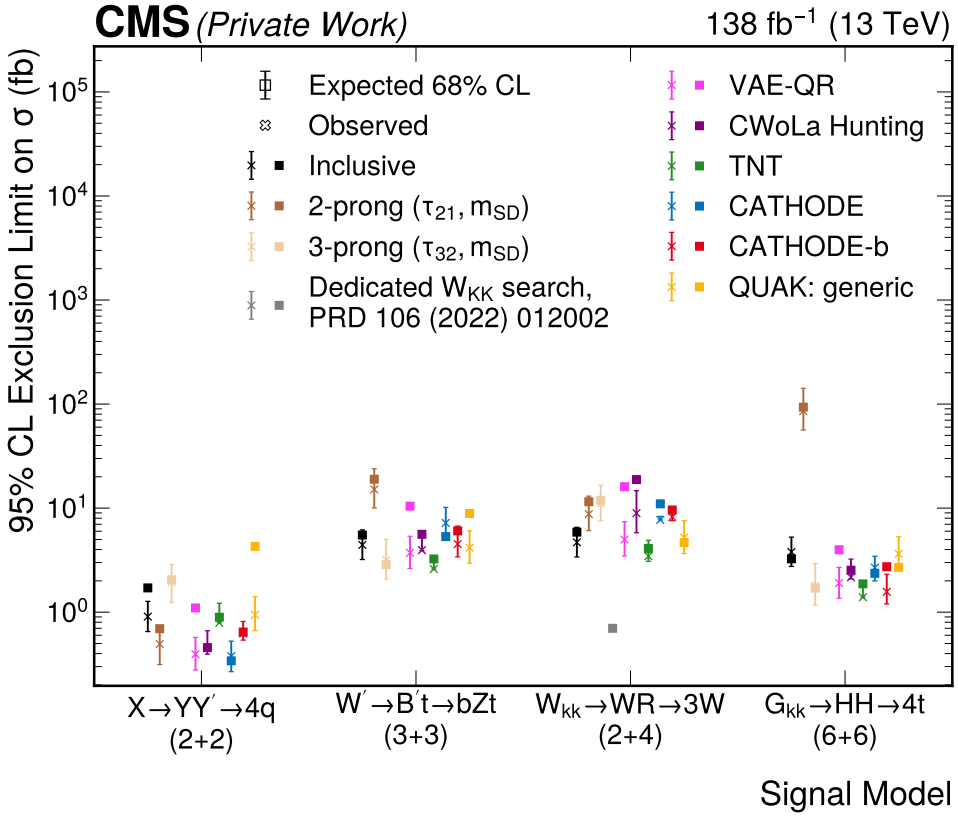


Figure 5.31.: Comparison of upper limits on the production cross section for benchmark signal models at 5 TeV, showing the relative performance of all five model-agnostic methods implemented in this analysis, alongside the inclusive approach.

Table 5.6.: Performance comparison of each anomaly detection method on different signal models of mass 3 TeV. The expected and observed limits are shown in parentheses (in units of fb), with the improvement factor quantifying the gain over the inclusive search. For example, a value of 1.7 implies an improvement of 70% over the inclusive search.

Signal Model	Method	Expected (Observed) limit	Improvement
$Q^* \to qW$ (25)	CWoLa	61.1 (30.1)	0.3
$Q^* \to qW$ (80)	CATHODE	46.2 (76.3)	0.4
$Q^* \to qW$ (170)	CATHODE	48.7 (86.3)	0.4
$Q^* \to qW$ (400)	CWoLa	45.8 (24.3)	0.5
$X \to YY' \to 4q$ (25/25)	CATHODE	7.4 (9.7)	0.9
$X \to YY' \to 4q$ (25/80)	CATHODE	5.9 (8.2)	1.2
$X \to YY' \to 4q$ (25/170)	CATHODE	8.3 (9.7)	0.8
$X \to YY' \to 4q$ (25/400)	VAE-QR	13.6 (12.5)	0.6
$X \to YY' \to 4q$ (80/80)	CATHODE	3.2 (4.3)	2.1
$X \to YY' \to 4q$ (80/170)	CATHODE	4.5 (6.0)	1.5
$X \to YY' \to 4q$ (80/400)	CATHODE	4.6 (6.0)	1.6
$X \to YY' \to 4q$ (170/170)	QUAK	2.7 (2.5)	2.6
$X \to YY' \to 4q$ (170/400)	CATHODE	4.3 (5.8)	1.7
$X \to YY' \to 4q$ (400/400)	VAE-QR	2.1 (1.9)	4.2
$W' \to B't \to bZt$ (25)	TNT	22.6 (13.9)	1.7
$W' \to B't \to bZt$ (80)	TNT	18.2 (11.3)	1.9
$W' \to B't \to bZt$ (170)	TNT	12.2 (7.3)	2.1
$W' \to B't \to bZt$ (400)	TNT	12.5 (7.0)	2.2
$W_{kk} \to WR \to 3W$ (170)	TNT	22.1 (15.2)	1.6
$W_{kk} \to WR \to 3W$ (400)	QUAK	19.7 (13.7)	1.8
$Z' \to T'T'$ (400)	TNT	39.1 (23.3)	3.5
$G_{kk} \to HH \to 4t$ (400)	VAE-QR	3.7 (3.2)	7.1

Part IV.

Quantum Machine Learning for High Energy Physics

6. Introduction to Quantum Mechanics and Quantum Computing

6.1. Fundamentals of Quantum Mechanics

Quantum mechanics is a fundamental theory of physics that describes the behaviour of matter and energy at atomic and subatomic scales. In contrast to classical mechanics, which describes phenomena deterministically, quantum mechanics integrates principles such as, but not limited to quantisation, wave-particle duality and superposition to create an inherently probabilistic description of reality.

At the heart of quantum mechanics lies the concept of a quantum state, which describes a physical system completely. Quantum states differ fundamentally from classical states since they allow superpositions, where a system can simultaneously exist in multiple possible configurations. Such states are mathematically represented by vectors in an abstract complex vector space known as the *Hilbert* space.

Closed quantum systems evolve according to the Schrödinger Equation, a deterministic unitary evolution preserving the quantum state's superposition properties. However, the process of measurement introduces fundamentally probabilistic outcomes, as described by the Born rule. This dichotomy leads therefore to the quantum measurement problem, leaving it an open question as to whether this the phenomenon of *wavefunction collapse* is physical or purely epistemological (see Section 6.1.2).

6.1.1. Dirac Notation and Inner Products

Quantum mechanics employs the Dirac, or bra-ket notation [173], which provides an elegant and powerful mathematical framework for describing quantum states. Within this formalism, quantum states are represented by vectors in a complex Hilbert space $\mathcal{H}$. The vectors themselves are denoted by *kets* $|\psi\rangle$, while their dual vectors are represented by *bras* $\langle\phi|$, belonging to the dual space $\mathcal{H}^*$.

An essential operation in Hilbert space is the inner product, defined for two vectors $|\psi\rangle$ and $|\phi\rangle$ as $\langle\phi|\psi\rangle$, yielding a complex number. This inner product satisfies the conjugate symmetry property:

$$\langle\phi|\psi\rangle = \overline{\langle\psi|\phi\rangle}, \tag{6.1}$$

and is positive semi-definite:

$$\langle\psi|\psi\rangle \geq 0, \tag{6.2}$$

with equality if and only if $|\psi\rangle$ is the zero vector. Given an orthonormal basis $|x\rangle$, any state vector $|\psi\rangle$ in $\mathcal{H}$ can be expressed as a linear combination of these basis vectors:

$$|\psi\rangle = \sum_x c_x |x\rangle, \quad \text{with coefficients } c_x = \langle x|\psi\rangle. \tag{6.3}$$

This representation simplifies the analysis of quantum systems, allowing efficient calculation of amplitudes and probabilities.

For composite quantum systems, Dirac notation conveniently expresses the tensor product of states. Given two individual systems in states $|\psi\rangle$ and $|\phi\rangle$, the state of the combined system is represented as:

$$|\psi\rangle \otimes |\phi\rangle \equiv |\psi\rangle|\phi\rangle, \tag{6.4}$$

highlighting the intuitive construction of multi-system quantum states from simpler constituents.

6.1.2. Quantum Measurements and Collapse

Quantum measurements fundamentally differ from classical measurements, as they exhibit inherently probabilistic outcomes and alter the state of the measured system. Measurement is mathematically described by a collection of measurement operators M_m, which must satisfy the completeness relation:

$$\sum_m M_m^\dagger M_m = I, \tag{6.5}$$

where I is the identity operator. When a measurement is performed on a quantum state $|\psi\rangle$, the probability of obtaining a specific measurement outcome m is given by:

$$p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle. \tag{6.6}$$

Subsequently, the quantum state collapses to a post-measurement state, described by:

$$|\psi_m\rangle = \frac{M_m|\psi\rangle}{\sqrt{p(m)}}. \tag{6.7}$$

A particularly important class of measurements, known as projective measurements, is characterised by measurement operators forming projections onto eigenstates of an observable. If the measurement basis is an orthonormal set $|x\rangle$, the measurement operators become $M_x = |x\rangle\langle x|$. The system then *collapses* from a superposition to a definite state $|x\rangle$ upon measurement outcome x, this uniquely quantum mechanical phenomenon appropriately being named as wavefunction collapse [174].

6.1.3. Hilbert Spaces and Composite Systems

Quantum states reside in a Hilbert space, and each distinct physical system is associated with a Hilbert space of dimensionality equal to its degrees of freedom. For instance, the simplest quantum system, which is the two-level qubit, has a two-dimensional Hilbert space $\mathbb{C}^2$. Here, the basis vectors commonly denoted as $|0\rangle, |1\rangle$ represent two distinct physical configurations. These could for example be the spin-up and spin-down states of a spin-$\frac{1}{2}$ particle.

When considering composite quantum systems composed of multiple subsystems, the corresponding joint state space is constructed using the tensor product of the individual subsystem Hilbert spaces. This mathematical construction provides a formal and systematic way to describe complex quantum systems built from simpler units. Specifically, for a system of n qubits, each individually described by the two-dimensional Hilbert space $\mathbb{C}^2$, the combined state space is given by the tensor product:

$$\mathcal{H}^{\otimes n} = \underbrace{\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2}_{n \text{ times}}. \tag{6.8}$$

This tensor product operation effectively combines the state spaces of individual qubits into a single larger Hilbert space. The dimension of the resulting composite Hilbert space scales exponentially with the number of qubits, explicitly given by:

$$\dim(\mathcal{H}^{\otimes n}) = 2^n. \tag{6.9}$$

The exponential growth in dimensionality is a hallmark feature of quantum systems, underscoring the complexity and computational potential of quantum mechanical systems. This lays the foundation for phenomena such as quantum entanglement, where composite states cannot be decomposed into independent subsystem states. The tensor product framework thus naturally accommodates entangled states, a crucial resource in quantum computing and quantum information processing, which will be extensively utilised and explored in subsequent sections.

6.1.4. Observables and Hermitian Operators

In quantum mechanics, physically measurable quantities such as position, momentum, energy, and spin are represented mathematically by operators acting on states within a Hilbert space. These operators are known as observables, and they must satisfy specific mathematical properties reflecting their physical nature. A key requirement for an operator to represent a physical observable is that it must be Hermitian (or equivalently, self-adjoint). Formally, an operator A is Hermitian if it satisfies the condition:

$$A = A^\dagger, \tag{6.10}$$

where $A^\dagger$ denotes the conjugate transpose (or adjoint) of the operator. Hermiticity ensures that the eigenvalues of the observable—interpreted physically as measurement outcomes—are always real numbers, consistent with physical intuition and experimental reality.

Every Hermitian operator can be expressed in its spectral decomposition form, which involves expanding the operator in terms of its eigenvalues and corresponding eigenstates. Mathematically, the spectral decomposition of a Hermitian operator A is expressed as:

$$A = \sum_i \lambda_i |a_i\rangle\langle a_i|, \quad \lambda_i \in \mathbb{R}, \tag{6.11}$$

where the eigenstates $|a_i\rangle$ form a complete orthonormal basis of the Hilbert space associated with the observable, and λ_i are the corresponding real eigenvalues.

When a quantum system described by a state vector $|\psi\rangle$ is measured with respect to an observable A, the expectation value (the statistically averaged measurement outcome over many repeated measurements of identically prepared states) is given by:

$$\langle A \rangle = \mathrm{Tr}(A|\psi\rangle\langle\psi|) = \sum_i \lambda_i |\langle a_i|\psi\rangle|^2. \tag{6.12}$$

This equation indicates that the probability of obtaining the eigenvalue λ_i in a measurement is $|\langle a_i|\psi\rangle|^2$, the squared magnitude of the projection of the state vector onto the eigenstate $|a_i\rangle$. Thus, the spectral decomposition explicitly encodes the probabilistic nature of quantum measurements, providing a direct connection between mathematical formalism and experimental outcomes.

For qubit systems specifically, a fundamental set of observables is given by the Pauli matrices $\{X, Y, Z\}$, which will be relevant throughout this part of the thesis. These three Hermitian matrices form a basis for all possible observables of a two-level quantum system (qubit), providing a complete representation of single-qubit measurements. In line with the convention used by major quantum circuit simulators, this thesis represents the Pauli matrices $\vec{\sigma}$ using the Latin alphabets X, Y, Z instead of $\sigma_x, \sigma_y, \sigma_z$. Their explicit forms are:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{6.13}$$

These matrices represent spin observables along the x, y, and z axes, respectively. As such, they play a crucial role in quantum information processing and quantum computing, serving as foundational building blocks for quantum gates and quantum circuits.

6.2. Introduction to Quantum Computing

6.2.1. Motivation and Basic Principles

Quantum computing has, over the past decade, emerged as a powerful computational paradigm capable of addressing problems beyond the reach of classical computation. Leveraging principles of quantum mechanics such as superposition and entanglement, quantum computers promise exponential speedups for certain computationally intensive tasks [175]. Notable examples include the Shor algorithm for integer factorisation [176] and the Grover Search algorithm for unstructured database searches [177], which demonstrate quadratic improvements over equivalent classical approaches. Additionally, quantum simulations of physical systems, inherently quantum in nature, can potentially surpass classical limits, enabling accurate modeling of molecular structures, chemical reactions, and particle physics interactions that are currently computationally prohibitive [178, 179].

In recent years, the intersection of quantum computing and machine learning has emerged as a promising avenue for enhancing performance and sensitivity in various domains, including HEP. Classical machine learning has, for some time, already been an essential tool in particle physics [81, 82], particularly within collider experiments such as the CMS and ATLAS Experiments. ML plays an important role in the pre-processing of immense quantities of data

recorded at the LHC, in addition to its application for tasks such as classification and anomaly detection. However, the exponential growth of data volumes and complexity at current and future colliders, such as the HL-LHC, poses significant computational challenges to traditional ML methods, where Quantum Machine Learning (QML) could potentially play a role.

QML presents an opportunity to overcome computational bottlenecks, potentially enabling faster processing and improved performance in data-intensive tasks. Specifically, quantum algorithms have demonstrated the capacity for efficient representation of high-dimensional feature spaces and faster training convergence, features particularly advantageous in HEP applications [180]. Two notable applications within particle physics are jet classification and anomaly detection. Given how crucial it is to perform jet classification accurately, it is plausible that in the near future, quantum [7] or hybrid quantum-classical algorithms could offer improved classification accuracy and efficiency by better capturing subtle correlations in the high-dimensional jet feature spaces compared to existing purely classical approaches.

Similarly, the identification of events that deviate significantly from known physics processes, or what is referred to as anomaly detection, is vital for discovering potential BSM signals. Classical anomaly detection techniques, while powerful, often face computational constraints and sensitivity limitations when applied to high-dimensional and complex data spaces, as is the case with collider events. Quantum algorithms such as quantum autoencoders [181] for example, have been explored in the past for the purpose of anomaly detection [8], leveraging their intrinsic ability to encode and process complex, high-dimensional data distributions efficiently.

Therefore, exploiting quantum computing and specifically QML methods in particle physics holds significant promise for addressing computational challenges posed by future HEP datasets, enhancing physics discovery potential, and paving the way towards new methodological paradigms in data analysis.

6.2.2. Qubits and Quantum State Representation

In classical computing a *bit* stores a single binary value, 0 or 1. A quantum computer replaces bits with *qubits*, which are two-level quantum systems whose pure states are elements of a two-dimensional Hilbert space [175]. A general qubit can be written as:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}, \tag{6.14}$$

$$\text{with } |\alpha|^2 + |\beta|^2 = 1, \tag{6.15}$$

where the complex coefficients (α, β) encode both probabilities along with a relative phase. Geometrically every qubit state corresponds to a point on the surface of the Bloch sphere; global phases are physically irrelevant.

Scalability. For a register of n qubits, the corresponding state vector lives in a 2^n-dimensional space,

$$|\Psi\rangle = \sum_{x\in\{0,1\}^n} c_x\ |x\rangle\,, \qquad \sum_x |c_x|^2 = 1,$$

with computational basis $|x\rangle = |x_1\rangle \otimes \cdots \otimes |x_n\rangle$. As shown in Equation 6.14, a single qubit can be represented as a column vector with 2 elements. Likewise, an n-qubit system can be represented as a column vector with 2^n elements. The representation for $n = 2$ is shown below:

$$|\psi_1\psi_2\rangle = \begin{pmatrix} \alpha_{00} \\ \alpha_{11} \\ \alpha_{10} \\ \alpha_{11} \end{pmatrix} \tag{6.16}$$

The basis states for the two-qubit system are $|00\rangle$, $|01\rangle$, $|10\rangle$ and $|11\rangle$, and their column vector representations are constructed by setting the corresponding element to one, and all others to zero. An example makes this abundantly clear as follows:

$$|10\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \tag{6.17}$$

It is standard practice to refer to the basis qubits of an n-qubit system as a state vector $|N\rangle \in \mathcal{H}^{\otimes n}$, with $N = 0, 1 \ldots 2^n - 1$. For example, the state $|2\rangle$ corresponds to the state $|10\rangle$ of Equation 6.17 for a two-qubit system.

This exponential growth in dimensionality potentially underlines a quantum speed-up but also poses serious challenges for simulations performed on classical computers in terms of compute power and memory requirements.

6.2.3. Quantum Gates and Circuits

Quantum dynamics must be norm preserving; hence logical operations are implemented by *unitary* matrices U that by definition satisfy the criterion $U^\dagger U = 1$. A small set of primitive one- and two-qubit gates can be shown to form a *universal* basis [182]. The gates that will be encountered most frequently in this thesis are defined below for easy reference:

- **Z-rotation:** This corresponds to a rotation about the Z axis on the Bloch sphere (described in Section 6.2.5). Its mathematical structure is given below, and one can likewise define the R_X and R_Y rotation gates.

$$R_z(\theta) = e^{-i\theta Z/2} = \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix}$$

- **Hadamard:** This physically corresponds to a rotation by an angle π about the axis defined by the direction $(\hat{x} + \hat{z})/2$, and creates the superposed states that conventionally describe the Hadamard basis.

$$H|0\rangle = |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \tag{6.18}$$

$$H|1\rangle = |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \tag{6.19}$$

Mathematically, this gate can be represented as:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

- **Controlled-NOT**: This is a two-qubit gate, crucial for creating entangled states. It requires two input qubits $|c; t\rangle$, the first being a control qubit and the second being the target qubit. Intuitively, one can visualise the CNOT gate as a controlled bit flip operation. If the control qubit is in the state $|0\rangle$, the target qubit is left unchanged. On the other, if the control qubit is in the state $|1\rangle$, then the state of the target qubit is flipped from $|0\rangle$ to $|1\rangle$, or vice-versa. As shown in Section 6.2.4, this gate can be used to create entangled states. Mathematically, it has the structure:

$$\text{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

  and its operation can be represented as:

$$\text{CNOT}\,|c; t\rangle = |c; c \oplus t\rangle \tag{6.20}$$

  where $\oplus$ is the Exclusive-OR (XOR) operation. The CNOT matrix acts on a two-qubit state vector of the form shown in Equation 6.17. One can easily perform this matrix multiplication to verify that the resultant two-qubit state is $\text{CNOT}\,|10\rangle = |11\rangle$.

A sequence of such gates forms a *quantum circuit*. The circuit *depth*, which is the longest sequence of gates acting successively on a single qubit, decides in most cases the runtime and on current Noisy Intermediate-Scale Quantum (NISQ) hardware, also the accumulated noise. Replacing fixed angles by variational parameters $\boldsymbol{\theta}$ produces what is called a Parameterised Quantum Circuit (PQC), setting the ground for QML in the scope of this thesis. Gradient-based optimisers [92, 183] running on a classical computer can then be used to iteratively update $\boldsymbol{\theta}$ to minimise a specific loss function.

6.2.4. Entangled States and Bell State Construction

For a bipartite system $\mathcal{H}_A \otimes \mathcal{H}_B$, a pure state is said to be *separable* if can be factorised as the tensor product of two states $|\psi\rangle = |a\rangle \otimes |b\rangle$. Otherwise the state is said to be *entangled*. Entanglement is a uniquely quantum phenomenon that ultimately enables applications such as quantum teleportation and super-dense coding [175]. One commonly encounters the two-qubit maximally entangled states, also known as the Bell States or Einstein-Podolsky-Rosen (EPR) Pairs, in quantum computing applications.

$$|\Phi^{\pm}\rangle = \frac{|00\rangle \pm |11\rangle}{\sqrt{2}}, \tag{6.21}$$

$$|\Psi^{\pm}\rangle = \frac{|01\rangle \pm |10\rangle}{\sqrt{2}}. \tag{6.22}$$

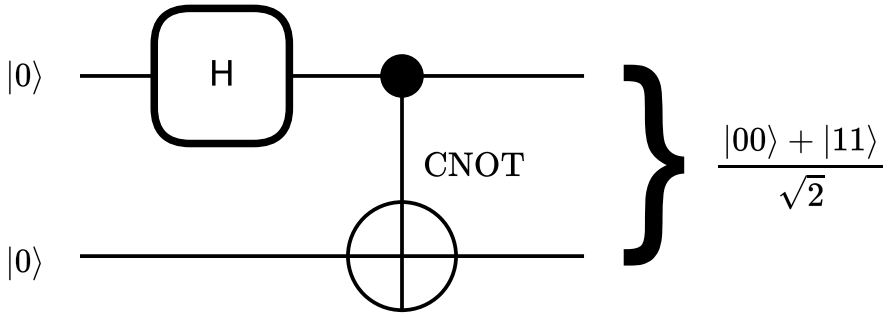


Figure 6.1.: A quantum circuit used to generate Bell States. The first operation is a Hadamard gate, whereas the second is the two-qubit CNOT gate of Equation 6.20, with the black dot denoting the control qubit, and the white circle the target qubit.

Starting from the $|00\rangle$ state, a circuit to create the $|\Phi^+\rangle$ Bell State is shown in Figure 6.1, with the following operations:

$$\Phi^+ = \text{CNOT}\ \left((H \otimes I)\,|00\rangle\right), \tag{6.23}$$

with the tensor product of operators in parentheses implying that the Hadamard gate is applied only to the first qubit of the two-qubit basis state $|00\rangle$.

6.2.5. Bloch Sphere Representation

The mathematical structure of a qubit allows one to visualise its two-dimensional Hilbert space representation on a three-dimensional unit sphere, this being the so-called *Bloch sphere*. Starting from the computational basis $\{|0\rangle, |1\rangle\}$, every normalised qubit state can be represented as shown in Equation 6.14. This can subsequently be reparametrised without loss of generality [175] as:

$$|\psi\rangle = \cos\frac{\theta}{2}\,|0\rangle + e^{-i\varphi}\sin\frac{\theta}{2}\,|1\rangle, \tag{6.24}$$

$$\text{where } \theta \in [0, \pi] \text{ and } \varphi \in [0, 2\pi).$$

It should be noted here that global phases are physically irrelevant. That being said, the two real parameters (θ, φ) describe a point on the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ (see Figure 6.2) with latitude $\frac{\pi}{2} - \theta$ and longitude φ. The north and south poles correspond to the basis-state kets $|0\rangle$ and $|1\rangle$ respectively, while $\theta = \pi/2$ traces out the equator, along which lie the states of equal superposition.

Pure and mixed states: In quantum mechanics a *pure state* is one which is precisely known, and can therefore be described by a single state vector $|\psi\rangle$ (up to an overall phase). Geometrically, for a single qubit a pure state corresponds to a point on the *surface* of the Bloch sphere. A *mixed state* on the other hand reflects classical ignorance about which pure state was actually prepared. It can be best pictured as an ensemble of several possible pure states $\{|\psi_k\rangle\}$, each

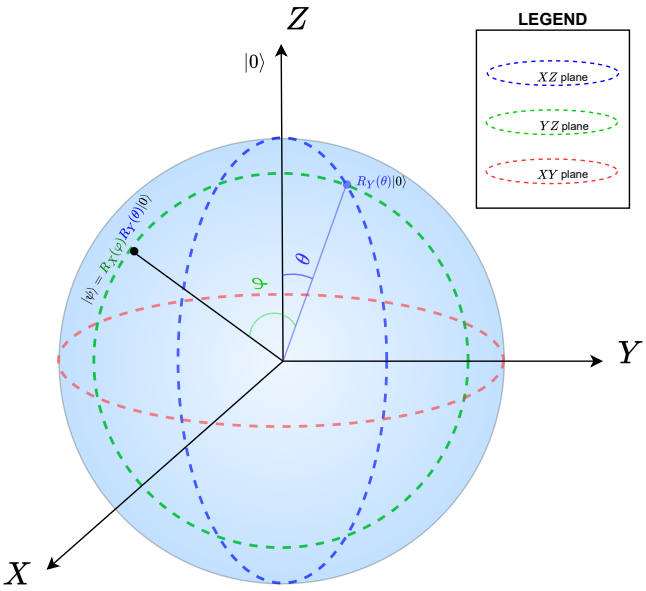


Figure 6.2.: A representation of a qubit state on the Bloch sphere, obtained by the successive application of an R_X and R_Y gate on the initial state qubit $|0\rangle$

chosen with some probability p_k. By convention, such states are represented using the notation $\{p_k, |\psi_k\rangle\}$.

Bloch vector and expectation values: Writing the Pauli matrices as $\vec{\sigma} = (X, Y, Z)$, one can define the *Bloch vector*

$$\vec{r} = (x, y, z) = \big(\langle X\rangle, \langle Y\rangle, \langle Z\rangle\big),$$

whose components in terms of α and β are

$$x = 2\,\mathrm{Re}\big(\alpha^*\beta\big), \qquad y = 2\,\mathrm{Im}\big(\alpha^*\beta\big), \qquad z = |\alpha|^2 - |\beta|^2. \tag{6.25}$$

For pure states $\|\vec{r}\| = 1$, and Eq. (6.24) maps (θ, ϕ) onto (x, y, z) via the transforms $x = \sin\theta\cos\phi, y = \sin\theta\sin\phi, z = \cos\theta$.

Density-matrix form: Both pure and mixed states can then be described by a 2×2 density operator

$$\rho = \frac{1 + \vec{r}\cdot\vec{\sigma}}{2}, \qquad 0 \le \|\vec{r}\| \le 1. \tag{6.26}$$

Pure states sit on the sphere's surface, whereas mixed states are located inside. One then defines the term purity as $\mathrm{Tr}(\rho^2) = (1 + \|\vec{r}\|^2)/2$, its value being 1 for pure states and smaller than 1 for mixed states.

Unitary operations as rigid rotations: An arbitrary single-qubit unitary operation can be expressed as:

$$U(\theta, \hat{n}) = e^{-\frac{i}{2}\theta\,\hat{n}\cdot\vec{\sigma}}, \tag{6.27}$$

This represents a rotation of the Bloch vector by an angle θ about the axis $\hat{n}$. The map

$$\Pi : \mathrm{SU}(2) \longrightarrow \mathrm{SO}(3), \tag{6.28}$$

is a *double cover* which means that two antipodal elements of SU(2) correspond to the same physical rotation in SO(3). Consequently, applying a 2π rotation in spin space, $U(2\pi, \hat{n}) = -1$, changes the global phase of $|\psi\rangle$ but leaves all observables invariant. It is only after a rotation of 4π that the state returns to itself.

The Bloch representation is therefore a powerful and intuitive framework for visualising quantum states and gate operations.

To conclude, Quantum Computing and QML hold the potential to revolutionise the landscape of HEP by offering novel paradigms for solving large-scale optimisation problems and increasing performance in tasks such as classification and anomaly detection. As quantum hardware moves beyond the NISQ era, these technologies may enable computational approaches that outperform current classical systems, particularly in the context of efficiently handling enormous data rates from future LHC upgrades, or developing interpretable algorithms for performing a specific task. The reader is encouraged to look at further comprehensive review papers [184, 185] that outline recent progress and summarise the application of quantum algorithms and QML models to key problems in HEP.

7. QML for HEP: Introducing 1P1Q

7.1. Introduction

Chapter 6 provided an overview of quantum mechanics and quantum computing. As established therein, quantum computing offers novel discovery paradigms that extend beyond the capabilities of current classical computational frameworks. In the domain of ML, significant attention has recently been directed towards Parameterised Quantum Circuits (PQCs), which serve as foundational building blocks for Quantum Neural Networks (QNNs). These quantum circuits, built up using parameterised unitary operations such as rotations, can be optimised in a manner analogous to classical NNs through variational approaches [186].

The application of quantum algorithms to HEP data represents an emerging field with potential advantages in both computational efficiency and representational capacity [187]. Specifically, quantum approaches to classification and anomaly detection have demonstrated promising preliminary results [7, 8, 188], and will also be the focus of this thesis. Quantum anomaly detection methods, particularly those based on Quantum Autoencoders (QAEs), leverage inherent quantum mechanical properties such as superposition and entanglement to potentially identify deviations in high-dimensional data that might elude classical techniques [8].

To evaluate the efficacy of QML algorithms for HEP applications, it is necessary to establish a systematic understanding of two elements:

- the mathematical structure and operation of PQCs.
- Quantum algorithms such as the QAE and the Variational Quantum Classifier (VQC)

This chapter addresses these aspects sequentially, providing an overview of both the theoretical foundation and the practical implementation.

Subsequently, the *One-Particle One-Qubit* (1P1Q) encoding method for mapping jet data onto quantum registers is introduced. This approach, which encodes the kinematic data of each individual jet PFCand on its own qubit, thereby representing a jet as a tensor product of Hilbert spaces, is demonstrated to be a particularly efficient encoding strategy. The results presented in this thesis establish the 1P1Q method as capable of delivering comparable performance across both anomaly detection and classification tasks, demonstrating its versatility as a quantum data representation method for HEP applications in the NISQ era.

It should be noted here that the current status of hardware in the NISQ era makes it computationally and financially unfeasible for the large circuits developed and trained in this thesis, to be deployed on a real quantum device. Consequently, all results presented in this thesis make use of a quantum simulator deployed on appropriate classical hardware, as will be mentioned in Section 7.7.3.

7.2. Quantum Circuits

The fundamental concept of the qubit, the basic unit of a two-level quantum system, was formally introduced in Section 6.2.2, while the mathematical framework of quantum operations was detailed in Section 6.2.3. These operations, commonly referred to as quantum gates, allow for the manipulation of quantum states through unitary transformations [175]. Each operation can be geometrically interpreted by visualising its action on the 3D Bloch sphere, as described in Section 6.2.5, where the state of a single qubit is visualised as a point on the surface of a three-dimensional unit sphere [182].

7.2.1. Introduction to Parameterised Quantum Circuits (PQCs)

A PQC can be formally defined as an ordered sequence of quantum gates, where certain gates depend on continuous parameters that can be adjusted [186]. Mathematically, a PQC with parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$ can be expressed as:

$$\mathcal{U}(\boldsymbol{\theta}) = \prod_{i=1}^{L} U_i(\theta_i)\,, \tag{7.1}$$

where each $U_i(\theta_i)$ represents a parameterised quantum gate, and the product indicates sequential application from right to left.

The constituent gates in a quantum circuit can be categorised into two primary types: single-qubit and multi-qubit operations. Single-qubit gates, such as the Pauli rotations ($R_X(\theta)$, $R_Y(\theta)$, $R_Z(\theta)$), act locally on individual qubits and perform rotations on the Bloch sphere around their respective axes by an angle θ. In contrast, multi-qubit gates operate on two or more qubits simultaneously and are essential for generating entanglement within the quantum system [182]. The best example of this is the CNOT gate of Equation (6.20).

The parameters $\boldsymbol{\theta}$ that define these parameterised gates can be optimised through either classical [92] or quantum-inspired [183] optimisation algorithms to achieve a specific objective [186]. In the context of QML, these objectives typically are tasks such as classification or anomaly detection. This optimisation process generally follows a variational approach, where parameters are iteratively adjusted based on measurement outcomes to improve performance.

A critical property of quantum circuits is *unitarity*, which ensures that the probability amplitude of the quantum state is preserved throughout the computation [189]. Mathematically, a matrix U is unitary if and only if:

$$UU^\dagger = U^\dagger U = I\,, \tag{7.2}$$

where $U^\dagger$ denotes the conjugate transpose of U and I is the identity matrix. This property guarantees that quantum operations are reversible and satisfy the conservation of probability required by quantum mechanics.

Overall, a quantum circuit may be viewed as a single composite unitary operation [175] which represents the cumulative effect of all gates in the circuit and transforms an initial quantum state $|\psi_{\text{in}}\rangle$ into a final state $|\psi_{\text{out}}\rangle$:

$$|\psi_{\text{out}}\rangle = \mathcal{U}(\boldsymbol{\theta})|\psi_{\text{in}}\rangle\,, \tag{7.3}$$

This mathematical formalism provides a foundation for implementing complex quantum algorithms through circuits made up of sequential unitary operations, with the true power of quantum computation emerging from properties such as superposition and entanglement that have no direct classical analogues.

7.2.2. Quantum Neural Network (QNN) Architecture

Having established the theoretical framework for quantum circuits, it is now possible to formally construct a QNN architecture, which is a variational quantum algorithm designed for machine learning. A prototypical QNN comprises the following essential components:

- **A register of n qubits that define the Hilbert space of the system**: These qubits are conventionally initialised in the computational basis state $|\mathbf{0}\rangle = |0\rangle^{\otimes n}$, which corresponds to the tensor product of n individual qubits in the $|0\rangle$ state. From a physical perspective, each $|0\rangle$ state represents an eigenstate of the Pauli-Z operator with eigenvalue +1, often interpreted as the spin-up configuration of a spin-$\frac{1}{2}$ particle [190].
- **Encoding of classical data into the quantum system**: Multiple data encoding strategies exist, with two prominent ones being angle and amplitude embedding [191, 192, 180]. The encoding operation $S(\mathbf{x})$ for input data $\mathbf{x} \in \mathbb{R}^d$ can be formally expressed as:

$$|\psi\rangle = S(\mathbf{x})\,|\mathbf{0}\rangle \tag{7.4}$$

- **Entanglement operations to establish correlations between qubits**: Entanglement, a uniquely quantum mechanical phenomenon with no classical analogue, is essential for capturing higher order relationships between input features [7, 186]. Without entanglement, the learning ability of the quantum circuit would be severely limited.
- **Trainable unitary operations $\mathbf{U}(\boldsymbol{\theta})$**: These operations are controlled by a vector of continuous parameters $\boldsymbol{\theta} \in \mathbb{R}^p$ that are systematically adjusted during the training process. The trainable unitary operation can be built in various ways, including layered architectures. The complete QNN can be formally expressed as:

$$|\psi_{\text{final}}(\mathbf{x}, \boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})\,|\psi_{\mathbf{x}}\rangle = U(\boldsymbol{\theta})S(\mathbf{x})\,|\mathbf{0}\rangle \tag{7.5}$$

- **Measurement**: The final component of a QNN involves measurement operations that project the quantum state onto a set of observables. This collapses the quantum state and extracts classical information that can be used to construct a cost function $\mathcal{L}(\boldsymbol{\theta})$. For a measurement operator M, the expectation value is calculated as:

$$\langle M\rangle_{\boldsymbol{\theta},\mathbf{x}} = \langle\psi_{\text{final}}(\mathbf{x}, \boldsymbol{\theta})|M|\psi_{\text{final}}(\mathbf{x}, \boldsymbol{\theta})\rangle \tag{7.6}$$

  The final measurement need not necessarily be performed on all qubits of the system, but could also be done on a subset of target qubits. This expectation value, sometimes combined with a classical post-processing step, is used to define a cost function that is minimised during the training process.

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) \tag{7.7}$$

In the next two sections, the algorithms used in this thesis to perform the tasks of anomaly detection and classification are described.

7.3. QAEs for State Compression

The classical autoencoder architecture was introduced and described in Section 4.6. To recap, an autoencoder is trained to compress the input information down to a latent space of lower dimensionality than the input, and subsequently reconstruct the input back again from this reduced representation. The goal is that the autoencoder, in this process, learns the underlying probability distribution that best describes its input dataset. Consequently, it performs poorly at reconstructing the inputs when they come from another probability distribution, and this reconstruction loss can be used as an anomaly metric.

To build a quantum circuit that implements the aforementioned principle is not trivial. Unlike classical operations where NN layers of different dimensionalities can be stacked up, no equivalent quantum operations can reduce dimensionality at will. This fundamental limitation arises from the unitarity constraint of quantum operations, which preserves the dimensionality of the Hilbert space [175]. The alternative therefore is to discard a subset of the input Hilbert space $\mathcal{H}_{\text{in}}$, by replacing it with the basis kets $|0\rangle$ from a separate *reference* Hilbert space $\mathcal{H}_{\text{ref}}$, a principle first introduced in Ref. [181].

7.3.1. Architectural Components

A quantum autoencoder [181] consists of the following architectural components:

- **Input Quantum Register:** The input qubits $|\Psi_{\text{in}}\rangle \in \mathcal{H}_{\text{IN}}$ of dimensionality N, with this being further subdivided into twp subsets designated as the latent space $\mathcal{H}_{\text{latent}}$ and the trash space $\mathcal{H}_{\text{trash}}$ such that $\mathcal{H}_{\text{IN}} = \mathcal{H}_{\text{trash}} \otimes \mathcal{H}_{\text{latent}}$. The total Hilbert space dimensionality is given by $\dim(\mathcal{H}_{\text{IN}}) = 2^n$ for an n-qubit system.
- **Encoder Circuit:** An encoder $\mathcal{U}_{\text{ENC}}(\boldsymbol{\theta})$ which is a QNN set up using the principles described in Section 7.2.2. This unitary transformation is parameterised by a set of variables $\vec{\boldsymbol{\theta}}$ that are optimised during training. The encoder acts on the input state according to:
$$|\Psi_{\text{encoded}}\rangle = \mathcal{U}_{\text{ENC}}(\vec{\boldsymbol{\theta}})\,|\Psi_{\text{in}}\rangle \tag{7.8}$$
- **Reference Qubits:** A set of reference qubits, initialised to $|0\rangle^{\otimes T} \in \mathcal{H}_{\text{REF}}$ or any other basis state. The dimensionality of this reference space must match that of the trash space: $\dim(\mathcal{H}_{\text{REF}}) = \dim(\mathcal{H}_{\text{trash}})$.
- **Decoder Circuit:** A decoder $\mathcal{U}_{\text{DEC}}(\vec{\boldsymbol{\phi}})$ that is allowed to act on the Hilbert space $\mathcal{H}_{\text{REF}} \otimes \mathcal{H}_{\text{latent}}$ to reconstruct the input as:
$$|\Psi_{\text{out}}\rangle = \mathcal{U}_{\text{DEC}}(\boldsymbol{\phi})(|0\rangle^{\otimes T} \otimes |\Psi_{\text{latent}}\rangle)\,, \tag{7.9}$$

where $|\Psi_{\text{latent}}\rangle \in \mathcal{H}_{\text{latent}}$ is the encoded state in the latent space. Mathematically, the decoder is the Hermitian adjoint $\mathcal{U}_{\text{DEC}}(\boldsymbol{\phi}) = \mathcal{U}^{\dagger}_{\text{ENC}}(\boldsymbol{\theta})$ of the encoder.

- **Loss Function:** An appropriately defined loss function to guide the training process. The quantum fidelity $F = |\langle\Psi_{\text{out}}|\Psi_{\text{in}}\rangle|^2$ is a natural choice of loss function $\mathcal{L}$, as it quantifies the similarity between the input and reconstructed quantum states. The training objective becomes:

$$\vec{\boldsymbol{\theta}}^*, \vec{\boldsymbol{\phi}}^* = \arg\max_{\vec{\boldsymbol{\theta}},\vec{\boldsymbol{\phi}}} F(|\Psi_{\text{in}}\rangle, |\Psi_{\text{out}}\rangle) \tag{7.10}$$

7.3.2. Training Procedure

The training of a quantum autoencoder involves the optimisation of parameters $\vec{\boldsymbol{\theta}}$ and potentially $\vec{\boldsymbol{\phi}}$ to maximise the fidelity between input and output states. However, as shown in Ref. [181], it is actually not necessary to train the decoder at all. Mathematically, the decoder needs to execute the operation of Equation 7.9, and since $\mathcal{U}_{\text{DEC}} = \mathcal{U}^{\dagger}_{\text{ENC}}$, this is exactly the same as training the encoder such that the trash states disentangle and go to the state $|0\rangle^{\otimes T}$. The set of trainable parameters that maximise the fidelity between the input and output qubit states, **are the same** as the ones that maximise the fidelity between the trash and reference qubit states. This is computationally far more efficient, and requires one to only train the encoder portion of the QAE circuit. With this in mind, the training procedure can be summarised as:

Algorithm 2 : Training a Quantum Autoencoder

1: Initialise the circuit parameters $\vec{\boldsymbol{\theta}}$ randomly, often constrained to lie in $[0, 2\pi]$ since they are rotation angles
2: **for** : each training input state $|\Psi_{\text{in}}\rangle$ from the dataset **do**
3: Prepare the input state on the quantum device
4: Apply the encoder $\mathcal{U}_{\text{ENC}}(\boldsymbol{\theta})$ unitary
5: Prepare the appropriate reference state $|0\rangle^{\otimes T}$
6: Measure the fidelity between the trash and reference states $|\langle\Psi_{\text{trash}}|\Psi_{\text{ref}}\rangle|^2 \in [0, 1]$
7: Compute the gradient of $-|\langle\Psi_{\text{trash}}|\Psi_{\text{ref}}\rangle|^2$ with respect to $\vec{\boldsymbol{\theta}}$, the minus sign being to maximise its value. Maximising this quantity is exactly equivalent to maximising the fidelity between the input and output states, as proved in Ref. [181].
8: **end for**
9: Update parameters using quantum-inspired methods such as quantum natural gradient [183] or a classical optimiser such as `ADAM` [92]
10: **Return** optimised encoder parameters $\vec{\boldsymbol{\theta}}$

A full schematic of the QAE is shown in Figure 7.1. As can be seen, the key metric here is the trash-reference fidelity, which quantifies how well the trash qubits are *reset* to the reference state [181]. Perfect compression would result in the trash qubits being completely disentangled and reset to the reference state with $F = 1$.

7.3.3. Measurement of Quantum Fidelity

Extracting the quantum fidelity, as seen in Step 6 of Algorithm 2 is not trivial. This is because it is not an observable, and therefore cannot be measured directly on a real quantum device,

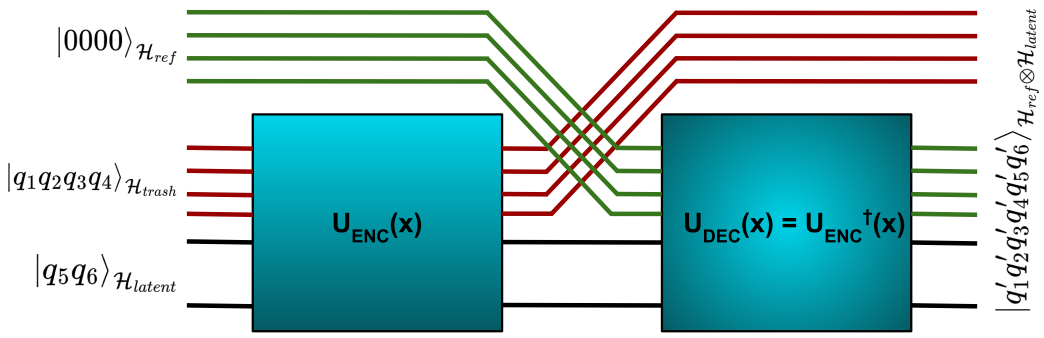


Figure 7.1.: Schematic for the QAE described in Section 7.3.1, with $\dim(\mathcal{H}_{\text{trash}}) = \dim(\mathcal{H}_{\text{ref}}) = 4$, $\dim(\mathcal{H}_{\text{latent}}) = 2$ and $\dim(\mathcal{H}_{\text{in}}) = 6$.

or inferred by measuring the states themselves. Instead, one must use what is known as the quantum SWAP test to obtain the quantum fidelity as the probability coefficient of one of the basis states describing the qubit, before subsequently extracting it through a measurement.

The quantum fidelity between two pure quantum states $|\psi\rangle$ and $|\phi\rangle$ is defined as the squared absolute value of their inner product:

$$F(|\psi\rangle, |\phi\rangle) = |\langle\psi|\phi\rangle|^2, \tag{7.11}$$

This quantity provides a measure of the similarity between two quantum states, with values ranging from 0 (orthogonal states) to 1 (identical states). In the context of quantum autoencoders, the fidelity between the trash and reference state(s) can be used as a loss function that quantifies the reconstruction quality [181].

The quantum SWAP test [193], is an elegant way to estimate the fidelity between two unknown quantum states. The circuit implementation requires three qubits: an ancillary qubit (or *ancilla*) initialized to $|0\rangle$, and two registers containing the states $|\psi\rangle$ and $|\phi\rangle$ whose fidelity is to be measured.

The test proceeds as follows:

1. Initialise the system in the state $|\psi_{\text{anc}}\rangle |\psi\rangle |\phi\rangle = |0\rangle |\psi\rangle |\phi\rangle$.
2. Apply a Hadamard gate to the ancilla qubit, resulting in:

$$\frac{1}{\sqrt{2}} (|0\rangle + |1\rangle) |\psi\rangle |\phi\rangle \tag{7.12}$$

3. Apply a controlled-SWAP operation (cSWAP), where the ancilla qubit controls whether the states $|\psi\rangle$ and $|\phi\rangle$ are to be swapped or not:

$$\frac{1}{\sqrt{2}} (|0\rangle |\psi\rangle |\phi\rangle + |1\rangle |\phi\rangle |\psi\rangle) \tag{7.13}$$

4. Apply another Hadamard gate to the ancilla qubit, resulting in:

$$\frac{1}{2}\big(\,|0\rangle\,(|\psi\rangle\,|\phi\rangle + |\phi\rangle\,|\psi\rangle) + |1\rangle\,(|\psi\rangle\,|\phi\rangle - |\phi\rangle\,|\psi\rangle)\big) \tag{7.14}$$

5. Measure the ancilla in the computational basis.

The probability of measuring the ancilla qubit in state $|0\rangle$ is given by:

$$\begin{aligned} P(0) &= \left\| \frac{1}{2}(|\psi\rangle\,|\phi\rangle + |\phi\rangle\,|\psi\rangle) \right\|^2 \\ &= \frac{1}{4}\,\| |\psi\rangle\,|\phi\rangle + |\phi\rangle\,|\psi\rangle \|^2 \\ &= \frac{1}{4}(2 + 2\mathrm{Re}(\langle\psi|\phi\rangle\,\langle\phi|\psi\rangle)) \\ &= \frac{1}{4}(2 + 2\mathrm{Re}(|\,\langle\psi|\phi\rangle\,|^2)) \\ &= \frac{1}{4}(2 + 2|\,\langle\psi|\phi\rangle\,|^2) \\ &= \frac{1}{2}(1 + |\,\langle\psi|\phi\rangle\,|^2) \\ &= \frac{1}{2}(1 + F(|\psi\rangle\,, |\phi\rangle)) \end{aligned} \tag{7.15}$$

Similarly, the probability of measuring the ancilla qubit in state $|1\rangle$ is:

$$P(1) = \frac{1}{2}(1 - F(|\psi\rangle\,, |\phi\rangle)) \tag{7.16}$$

Thus, the fidelity between the states can be calculated as:

$$F(|\psi\rangle\,,\, |\phi\rangle) = 2P(0) - 1 = 1 - 2P(1) \tag{7.17}$$

This result demonstrates how the quantum `SWAP` test (with the circuit shown in Figure 7.2) elegantly transforms the fidelity into a measurable quantity through a simple quantum circuit. For completeness, it should be noted that the statistical nature of quantum measurements means that multiple repetitions of the `SWAP` test are required to estimate the fidelity with reasonable precision. The variance of this estimate scales as $O(1/\sqrt{M})$, where M is the number of measurement repetitions (commonly referred to as **shots**), following standard sampling statistics.

7.3.4. Anomaly Detection with QAEs

The application of QAEs to anomaly detection in HEP follows the same conceptual framework as classical autoencoders with the reconstruction fidelity serving as the anomaly metric.

Potential advantages of QAEs for anomaly detection include their ability to handle inherently quantum data that contains complex correlations, in addition to capturing features that classical autoencoders might miss. Additionally, the intrinsic unitarity of quantum operations leads to a

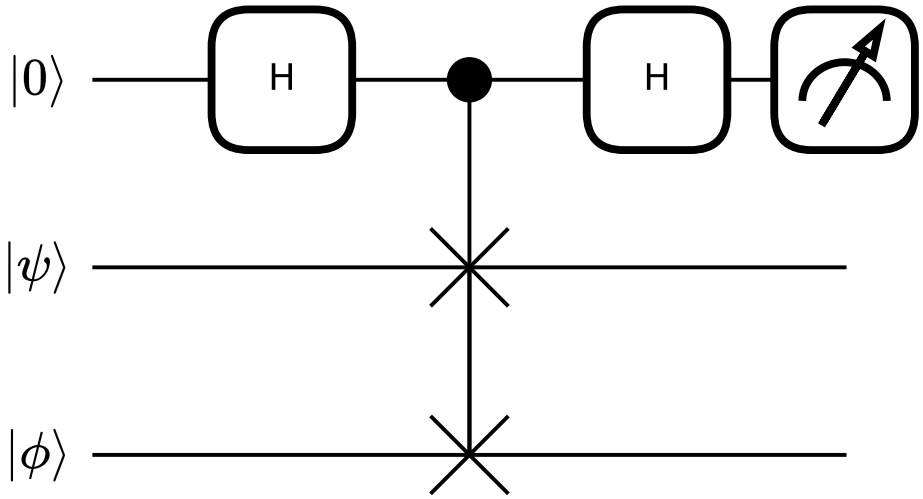


Figure 7.2.: Circuit to perform the quantum SWAP test described in Section 7.3.3. The initial Hadamard gate H is followed by the cSWAP operation, with the black dot representing the control, or ancillary qubit and the two crosses denoting the qubit states (in this case, the trash and reference states) which are to be swapped based on the state of the ancillary qubit. This is followed by the application of another Hadamard gate, and finally a measurement of the Z operator expectation value on the ancillary qubit, denoted by the needle gauge, to extract the value of the quantum fidelity between the two states $|\psi\rangle$ and $|\phi\rangle$ (see Equations 7.17 and 7.30)

different inductive bias when compared to classical NNs, which might prove to be advantageous for certain classes of problems [194].

In the context of HEP, QAEs have been successfully applied to jet classification and anomaly detection tasks, where the quantum circuit learns to distinguish between different jet types based on their constituent particle properties, or to identify when the jet differs significantly from the background [7, 8]. These implementations demonstrate the potential of quantum algorithms as the hardware continues to advance beyond the current NISQ era.

7.4. Classification on Quantum Computers: the Variational Quantum Classifier (VQC)

Section 7.2.2 introduced the formalism of a QNN. This section quickly introduces how they can be used to perform the task of classification, for example on jets. The basic idea is the same as was described for a QAE, in Section 7.3.1. Data is encoded using operations such as rotations onto a qubit system. This is followed up by a series of entanglement operations that ensures the network has enough flexibility to share information between qubits and learn higher order correlations from the features. Thereafter, trainable parameterised unitary operations are applied, to allow the QNN to reach a state that minimises the given objective. The final prediction from the QNN is extracted in the form of a measurement of an observable, usually on one or a set of *target* qubits. This predicted label is finally optimised to be as close as possible to the true label of the training input.

The main difference between a QAE and a VQC lies in the training objective, as is also the case classically. While autoencoders work to minimise the reconstruction loss between the input

and reconstructed output, a classifier works to minimise an appropriate defined loss between a predicted label from the network and the true label. The main difference lies in the learning paradigm, with autoencoders being unsupervised learners that learn to recognise anomalous patterns in data without labels, whereas classifiers follow the supervised learning paradigm, requiring the presence of clearly labeled data for training and validation.

In a typical VQC architecture, a quantum state $|\psi_{\text{in}}\rangle$ representing the input data is prepared through a feature map or encoding circuit $U_\phi(x)$. This is followed by a variational quantum circuit $U(\theta)$ with trainable parameters θ. The circuit can be expressed as:

$$|\psi_{\text{out}}\rangle = U(\theta)U_\phi(x)\,|0\rangle^{\otimes n}\,. \tag{7.18}$$

where $|0\rangle^{\otimes n}$ represents the initial state of n qubits. The measurement of specific qubits then provides the classification output. For a binary classification task, the expectation value of a measurement operator $\hat{M}$ (often the Pauli-Z operator on a designated output qubit) is used:

$$\langle\hat{M}\rangle = \langle\psi_{\text{out}}|\hat{M}|\psi_{\text{out}}\rangle. \tag{7.19}$$

The resulting expectation value $\langle\hat{M}\rangle$ typically lies in the range $[-1, 1]$, which can be rescaled to $[0, 1]$ for interpretation as a probability via the transform $p = \frac{\langle\hat{M}\rangle+1}{2}$.

A crucial aspect of VQC design is the circuit architecture, which must balance expressivity with trainability. The layered ansatz approach is sometimes employed, consisting of alternating data encoding and variational layers [186, 195]. This architecture is particularly suited for the NISQ era, where circuit depth is limited by noise and decoherence effects [196].

The most common loss function for training a binary classifier is the Binary Cross Entropy (BCE):

$$L_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right] \tag{7.20}$$

where N is the number of training samples, y_i is the true label (0 or 1) for the ith sample, and $\hat{y}_i$ is the predicted probability from the model for the i^{th} sample belonging to the class 1.

Another choice, although less common is the Mean-Squared Error (MSE):

$$L_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{7.21}$$

The optimisation of parameters θ presents unique challenges on quantum devices. One alternative is the usage of classical optimisers, typically in a hybrid quantum-classical approach. Gradient-based methods such as stochastic gradient descent or Adam can be implemented using parameter-shift rules to compute gradients on quantum hardware [183].

Key challenges in implementing effective VQCs include the barren plateau problem [197], whereby gradients vanish exponentially with increasing circuit size, and the robustness of the classifier against noise. Various mitigation strategies have been proposed, including

circuit structure optimisation and error mitigation protocols [186]. However, most current implementations remain as proof-of-concept only due to hardware limitations, although they do still demonstrate a potential advantage in capturing complex correlations that might be difficult for classical models to suitably learn.

7.5. The 1P1Q Data Encoding Scheme

The application of QML techniques to jet physics presents unique challenges due to the high-dimensional nature of jet data and the limited number of qubits available in current NISQ devices [187]. Existing approaches that attempt to use QML for jet classification or anomaly detection have predominantly employed two strategies:

- Using pre-computed jet-level features [7, 8, 198], which may discard potentially valuable information contained in the raw data, or
- Implementing hybrid quantum-classical pipelines where classical NNs first compress the input features before feeding them into a QNN [199].

The fundamental challenge in quantum data encoding and QML lies in efficiently mapping classical data to the corresponding quantum states. Standard encoding methods require increasingly many qubits as the feature dimension increases, while often failing to capture complex correlations between features. These limitations become prominent when dealing with jet physics data, where the number of constituents and their associated features can vary significantly.

7.5.1. Outline

In the forthcoming sections, the 1P1Q encoding scheme is formally introduced, and subsequently evaluated using two quantum algorithms: a QAE for unsupervised anomaly detection (Section 7.3) and a VQC for supervised classification (Section 7.4). A full description of the datasets used for the training purpose, along with the details of the training process, is provided. Finally, the results obtained for the two aforementioned tasks are presented. In addition, benchmarks are set against equivalent and state-of-the-art classical algorithms, and the results from studies performed to understand and interpret the performance of the 1P1Q approach are presented.

Through this part of the thesis, the potential of quantum computing to address challenging problems in jet physics is examined, and it is shown that the 1P1Q approach can offer advantages in terms of accuracy, interpretability, and computational efficiency compared to existing methods.

7.5.2. Advantages over Classical and Hybrid Approaches

An encoding like 1P1Q offers several key advantages over traditional approaches:

First, by using constituent-level features directly, the scheme avoids any potential information loss that occurs when calculating jet-level observables, which are typically derived through

mathematical operations that may not capture all relevant correlations in the data. This is particularly important for anomaly detection, where unexpected features might be precisely what distinguishes signal from background.

Second, unlike hybrid approaches that employ classical NNs for dimensionality reduction before passing it on to the QML algorithm, 1P1Q preserves the full information content of the input data. Classical compression steps, while computationally efficient, may inadvertently discard subtle patterns that could be crucial for distinguishing rare signals. This is similar to the point made above, except that a classical NN is allowed to learn the best possible discriminating variables, which can then be used as inputs to the QNN.

Third, 1P1Q offers superior interpretability and far lower parameter complexity compared to black-box NN approaches. Since each qubit corresponds to a specific jet constituent, the quantum circuit's behaviour can be more readily understood, with each unitary transform allowing one to visualise how the quantum state initially representing the particle changes. Additionally, a quantum algorithm using such an encoding requires far fewer parameters compared to a classical model that performs the same task, often by more than 2 orders of magnitude. With far fewer parameters and easily visualisable intermediate states, one can use this to gain insights into which constituent particles or properties, or which quantum operations significantly influence the learning process.

7.5.3. The 1P1Q Approach

The 1P1Q data encoding scheme [200] introduced in this thesis represents a novel approach designed to address these challenges. As the name suggests, 1P1Q encodes each jet constituent particle onto its *own* qubit, establishing a direct correspondence between the physical particles and their quantum representation. This approach differs fundamentally from existing methodologies by:

- Preserving the constituent-level information without loss through an intermediate transformation or compression step
- Facilitating interpretability by maintaining a one-to-one mapping between particles and qubits

In the 1P1Q scheme, each jet constituent's kinematic properties are encoded into the quantum state of a single qubit through a systematic application of rotation gates. The constituent-level features employed in this encoding are the normalised transverse momentum ($p_T/p_{T,\text{jet}}$), relative pseudorapidity ($\eta_{\text{rel}} = \eta - \eta_{\text{jet}}$), and relative azimuthal angle ($\phi_{\text{rel}} = \phi - \phi_{\text{jet}}$), the latter two quantities being calculated relative to the jet axis. These fundamental parameters are selected specifically because they encapsulate the complete kinematic information of each constituent, from which all higher-level jet features can be subsequently derived. For each constituent particle within the jet, this information is encoded into one qubit (hence the nomenclature *1P1Q*) as follows:

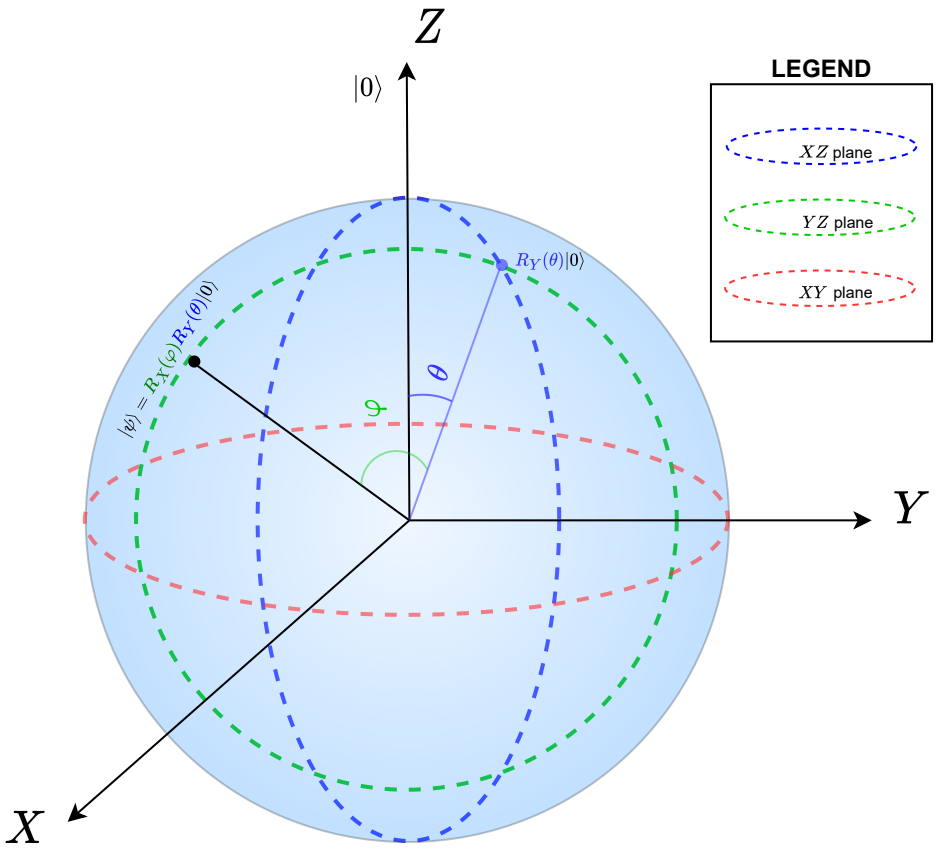


Figure 7.3.: The state of a single input qubit corresponding to one jet constituent particle, shown after the first R_Y rotation (blue) and the second R_X rotation (green), leading to the final input state (black) described in Equation 7.26. The planes in which these rotations occur are highlighted for a clearer understanding.

$$f \cdot \frac{p_\mathrm{T}}{p_{\mathrm{T,jet}}} \cdot (\eta - \eta_\mathrm{jet}) \rightarrow \theta\,, \tag{7.22}$$

$$f \cdot \frac{p_\mathrm{T}}{p_{\mathrm{T,jet}}} \cdot (\phi - \phi_\mathrm{jet}) \rightarrow \varphi\,, \tag{7.23}$$

$$\text{where} \quad f = 1 + \frac{2\pi}{1 + e^{-w}}\,, \tag{7.24}$$

Subsequently, the angles (θ, φ) are utilised as parameters for the unitary rotations R_X and R_Y:

$$\left(\frac{p_\mathrm{T}}{p_{\mathrm{T,jet}}}, \eta_\mathrm{rel}, \phi_\mathrm{rel}\right) \rightarrow |\psi\rangle_j = R_X(\varphi) R_Y(\theta)\,|0\rangle_j\,, \tag{7.25}$$

The operations used to construct the 1P1Q encoding are shown in Figure 7.3.

The quantities η_rel and ϕ_rel are typically constrained within the range $[-0.8, 0.8]$ for AK8 jets, and are subsequently rescaled to the range $[-\pi, \pi]$. This is because they function principally as rotation angles whose inherent periodicity can be exploited. An additional trainable parameter $f(w)$, constrained to the interval $[1, 2\pi + 1]$, is introduced to scale the inputs. This scaling

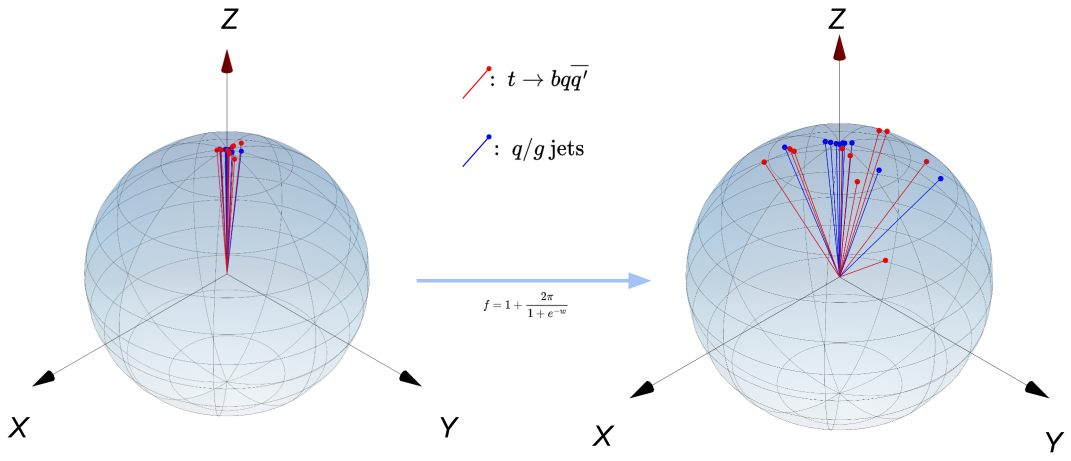


Figure 7.4.: Effect of the input scaling factor f on two jets, one initiated by a light quark/gluon (blue) and one by a top quark (red).

factor was determined to be both necessary and effective for the specific implementation of the VQC, primarily due to the small values of $p_T/p_{T,\text{jet}}$ causing jet constituents to cluster in proximity to the North Pole of the Bloch sphere. The effect of this trainable scaling factor is illustrated in Figure 7.4, which demonstrates that for two jets possessing comparable momentum, one initiated by a top quark and the other by a light quark/gluon, the constituent representations become more distinctly separated on the Bloch sphere, thereby enhancing the discriminative capability of the VQC. In the particular case of the model-agnostic QAE, which is trained exclusively on background jets, the trainable scale factor was observed to converge to a value approximating unity, effectively nullifying its influence.

Each input qubit corresponding to a single jet constituent can, following the 1P1Q encoding procedure, be represented mathematically as:

$$\begin{aligned}(p_T/p_{T,\text{jet}}, \eta_{\text{rel}}, \phi_{\text{rel}}) \to |\psi\rangle = &\left(\cos\frac{\theta}{2}\cos\frac{\varphi}{2} + i\left(\sin\frac{\theta}{2}\sin\frac{\varphi}{2}\right)\right)|0\rangle \\ &+ \left(\sin\frac{\theta}{2}\cos\frac{\varphi}{2} - i\left(\cos\frac{\theta}{2}\sin\frac{\varphi}{2}\right)\right)|1\rangle\end{aligned} \tag{7.26}$$

The complete input space after the encoding process is represented by the tensor product $\bigotimes_{j=1}^{N} |\psi\rangle_j$, where N denotes the number of input particles considered per jet. Hardware constraints impose restrictions on how high this number can be, since the memory and computational requirements for a quantum simulator scale exponentially with the number of qubits present in the system.

7.5.4. Model Bias

An important consideration in the 1P1Q approach is to have inherent robustness against bias introduced by the scale of the event and of individual jet constituents. The absolute kinematic characteristics of jets originating from light quarks or gluons, which are fully described within the framework of the SM, are expected to differ substantially from those containing exotic BSM signatures, or even from jets initiated by SM processes such as the top quark decay. Consequently, any effective method must demonstrate invariance to the absolute value of the jet's energy or momentum, as well as to the corresponding properties for individual jet constituents. Furthermore, the presence of distinctive features in kinematic distributions, such as peaks in the soft-drop mass m_{SD}, must be appropriately accounted for or decorrelated from the anomaly metric, as was done for the VAE-QR methodology discussed in Chapter 5.

This bias mitigation is systematically accomplished through two steps:

- As formulated in Equation 7.22, only the relative p_{T} carried by an individual jet constituent, expressed as a fraction of the total jet p_{T}, is encoded into the qubit state. This approach effectively prevents the QNN from being influenced by the presence of high-momentum individual jet constituents, which naturally occur with greater frequency in energetic jets, irrespective of their actual origin.
- The kinematic distributions characteristic of jets arising from decays such as $t \to bq\overline{q'}$ or $\mathrm{Z} \to q\overline{q'}$, are inherently distinct from those of jets initiated by light quarks or gluons. To ensure that the QNN remains unbiased towards the presence of peaks in kinematic distributions, or more generally, to preclude the network from associating higher energy signatures with anomalous or signal-like behaviour, jets utilised in the training, validation, and testing datasets are specifically sampled to have a uniform p_{T} distribution. This sampling methodology ensures that the QNN does not erroneously correlate higher momentum with higher signal probability, even if this might often be the case. The distributions before and after this sampling are shown in Figure 7.5.

7.6. Jet Datasets

Datasets that have been used in the past for machine learning on jets include:

- The JetNet [201] dataset which contains top-, light- and other heavy quark-initiated jets,
- The Top Tagging [202] dataset which contains jets from a top quark decay, and those initiated by light quarks or gluons, and
- The Quark-Gluon [203] dataset containing jets initiated either by a light quark, or gluon.

7.6.1. The JetClass dataset

The JetClass dataset [82] is more recent and far larger in scope and size compared to any of its predecessor datasets used within the HEP community, and is therefore used for the purpose

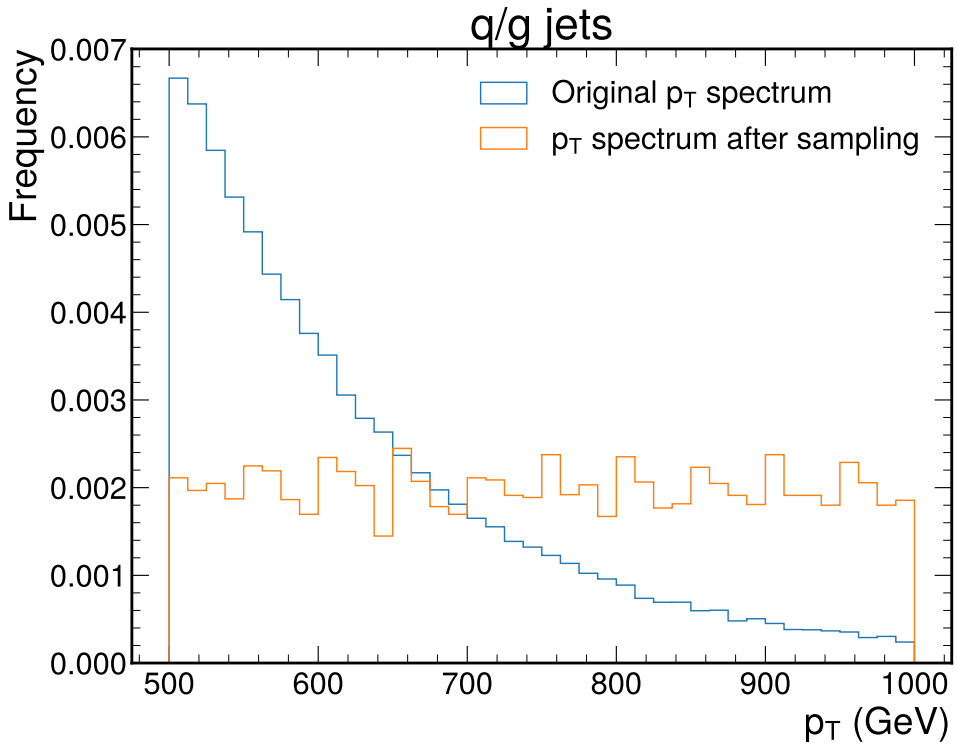


Figure 7.5.: Effect of the p_T sampling on the input datasets used for training the QAE and the VQC

of this thesis to train and benchmark the 1P1Q approach. A short description is provided below.

This dataset contains a total of 100M jets divided into 10 classes. The background jets are those initiated by light quarks or gluons, while jets arising from decays such as, but not limited to, $t \to bq\overline{q'}$ and $W/Z \to q\overline{q'}$ are treated as signal jets. The production and subsequent decay of top quarks and W, Z, and Higgs bosons are simulated with MADGRAPH5_aMC@NLO at NLO precision [120]. Parton showering and hadronisation processes are simulated in PYTHIA [117].

To ensure the simulated jets closely resemble those reconstructed by the ATLAS or CMS detectors, these detector effects are simulated with DELPHES [204] using a simplified CMS detector configuration. Jets are clustered utilising the anti-k_T algorithm [39] with a distance parameter $R = 0.8$. As additional criteria, only jets with $p_T \in [500, 1000]$ GeV and pseudorapidity $|\eta| < 2$ are stored. For signal jets, additional quality requirements are imposed to ensure they fully contain the decay products of the initial particles.

7.6.2. The Aspen Open Jets (AOJ) Dataset: Moving to real data

The CMS Collaboration has previously released authentic collision data recorded during the Run 2 (2016-18) data-taking period. The Aspen Open Jets (AOJ) dataset [205] is derived from the CMS 2016 JetHT datasets [206, 207] and presented in a structured format specifically optimised for machine learning applications. Although an extensive description of CMS data acquisition and processing falls beyond the scope of this thesis, it is important to note that this

dataset predominantly comprises jets initiated by light quarks or gluons, with contamination from alternative decay processes such as those involving W/Z bosons or top quarks being less than 1% of the total sample size. In its entirety, the dataset contains approximately 180M jets recorded in 2016 with the CMS detector. Within the scope of this thesis, this dataset is used to prove that the 1P1Q approach is able to learn the underlying physics of a dataset without being biased by artefacts of the detector simulation process.

7.7. Training Setup for 1P1Q: Datasets and Circuits

As mentioned in Section 7.5.1, the 1P1Q methodology is used to perform the two tasks of anomaly detection and supervised classification. In each case, the inputs to the circuit are the N highest p_T jet constituents, since these are expected to contain the most information. Due to computational constraints, the largest circuit used in this thesis contains $N = 10$ input qubits.

7.7.1. QAEs for Anomaly Detection

The QAE was formally introduced in Section 7.3. In this section, only the specifics of the circuit used are described. As mentioned above, the input space consists of N qubits. Following the encoding operations of Equation 7.25, the QAE, or rather its encoder (see Section 7.3.1) is constructed as follows:

- Entanglement operations using `CNOT` gates with each possible pair of qubits in the input space being used as the control and target respectively for the `CNOT` gate. This requires a total of $\frac{1}{2}N(N-1)$ operations.
- 3 trainable rotation gates per qubit, these being the R_Y, R_Z and R_X rotations respectively, for a total of $3N$ operations (and therefore trainable parameters)

The unitary operation describing the encoder of the QAE used in this thesis is shown in Figure 7.6a and can be written down as:

$$U_{\text{ENC}}(\Theta) = \left(\bigotimes_{i=1}^{N} R_X(\delta_i) R_Z(\gamma_i) R_Y(\omega_i) \right) \otimes \left(\bigotimes_{1 \le i < j \le N} C_{ij} \right), \tag{7.27}$$

where $(\delta_i, \gamma_i, \omega_i)$ is the vector of trainable parameters associated with qubit i.

As also described in Section 7.3.2, it is not actually necessary to train both the encoder and decoder components of the QAE. In an equivalent approach, the trash state qubits are optimised to converge to the $|0\rangle$ reference states. The corresponding loss function is the negative of the Quantum Fidelity of Equation 7.11, which is calculated using the Quantum `SWAP` test described in Section 7.3.3. The choice of trash, latent and reference space dimensions are hyperparameters that must be optimised to strike the ideal balance between too much surviving information, and too little. These are subject to the obvious constraints:

$$N_{\text{trash}} + N_{\text{latent}} = N \tag{7.28}$$

$$N_{\text{trash}} = N_{\text{ref}} \tag{7.29}$$

For an input of N qubits, out of which N_{trash} are discarded, the total number of qubits required in the QAE circuit is $N + N_{trash} + 1$, the last qubit being the ancilla used to perform the `SWAP` test. For $N = 10$ qubits and a latent space of only 1, the largest circuit therefore simulated within the scope of this thesis contained 20 qubits. A smaller circuit containing 4 inputs and a total of 8 qubits is shown in Figure 7.6a

Equations 7.15 and 7.16 show the probability of the ancilla qubit being in either basis state after performing the `SWAP` test. The final fidelity can be extracted by performing a measurement of the expectation value of the Pauli-Z operator (corresponding to the Z matrix) on the ancilla qubit, which proceeds as:

$$\begin{aligned} \langle \psi_{\text{anc}} | \, Z \, | \psi_{\text{anc}} \rangle &= P(0)^2 - P(1)^2 \\ &= F(|\psi\rangle, |\phi\rangle), \end{aligned} \tag{7.30}$$

The QNN is trained to optimise the negative of the quantity shown in Equation 7.30.

7.7.2. VQCs for Supervised Classification

As described in Section 7.4, the VQC is a QNN, similar in structure to the encoder of the QAE. For an input space consisting of N qubits, encoded using the 1P1Q scheme, the following elements make up the VQC circuit used in this thesis:

- Entanglement operations using `CNOT` gates on each adjacently numbered pair of qubits, leading to a total of N operations. This was observed to have the best performance for supervised classification, with the usage of entanglement operations on each possible pair of qubits leading to a slightly worse performance overall.
- As was the case for the QAE, three parameterised rotations are applied to each qubit, for a total of $3N$ operations.

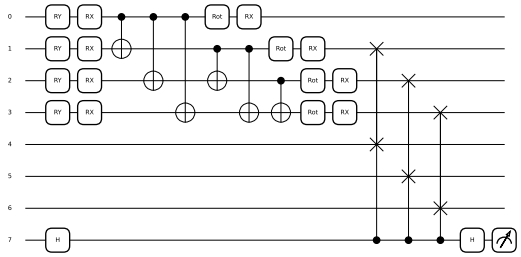
The unitary operation describing the VQC can be summarised as:

$$\begin{aligned} U(\Theta) = &\left(\bigotimes_{i=1}^{N-1} R_X(\delta_i) R_Z(\gamma_i) R_Y(\omega_i) \right) \\ &\otimes \left(\bigotimes_{i=0}^{N} C_{i,(i+1) \bmod N} \right), \end{aligned} \tag{7.31}$$

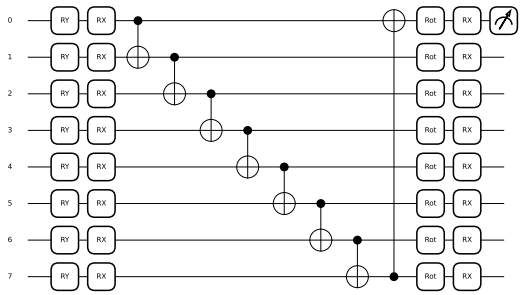
Unlike the QAE, the VQC is trained to separate signal from background, and therefore requires clearly labeled data. These labels y_i are then compared to the predictions $\hat{y}_i$ from the VQC using the MSE loss function (see Equation 7.21), with the addition of an extra classical bias term to the prediction from the network, which has been shown to smoothen the output [7]. Once again, it is ultimately the parameters $(\delta_i, \gamma_i, \omega_i)$, with $i = 1 \ldots N$ that need to optimised to fulfill the classification objective.

$$\hat{y} = \langle \psi_{\text{target}} | \, Z \, | \psi_{\text{target}} \rangle + b, \tag{7.32}$$

where $|\psi_{\text{target}}\rangle$ is the target qubit, which in the scope of the thesis is the qubit 0 onto which the hardest jet constituent is encoded. The VQC circuit with N inputs (shown in Figure 7.6b for $N = 8$) contains a total of $3N + 2$ trainable parameters.



(a) The QAE circuit described in Equation 7.27 for an input of 4 particles, and a latent space of 1 qubit.



(b) The VQC circuit described in Equation 7.31 is shown here for an input space of 8 qubits.

7.7.3. Datasets and Training Setup

The datasets used in the scope of this thesis have already been described in Section 7.6. The specific points regarding each algorithm are as follows:

- **QAE**: This is trained on up to 10000 jets from the simulated JetClass dataset, each such jet being initiated by a light quark or gluon. Being an unsupervised learner, the QAE does not require signal for its training. Its performance is then evaluated on the basis of its ability to distinguish such light quark or gluon jets from signal jets that are typically expected to have complex substructure and higher pronginess. Thereafter, the QAE is also trained on the same number of jets from the `AOJ` dataset (which contains real data), and evaluated using signal models from the JetClass dataset (see Section 7.8.1). For the purpose of evaluation, 10000 jets of each type are used.
- **VQC:** This is trained on up to 1000 jets from the simulated JetClass dataset as well, equally subdivided into two classes, one being jets initiated by a light quark or gluon, and the other being the jets arising from the decay of a top quark $t \to bq\overline{q}'$. Effectively, the VQC is a top tagger. Once again, for the purpose of evaluation, 10000 jets of each type are used.

In each case, the jets used for training, validation and evaluation are sampled (see Section 7.5.4 for the motivation) to have a flat p_{T} spectrum in the range $(500, 1000)$ GeV, carried out in bins of width 50 GeV, as shown in Figure 7.5.

The circuits are executed using the quantum simulator library Pennylane [208] in `Python`. Gradient computation and optimisation of parameters is carried out using the Autograd library [209] and the Adam optimiser.

7.8. Results

The results obtained using the 1P1Q method are currently available as a preprint [200] on the `arXiv` repository. Several figures presented in this section have been adapted from this preprint. For the purpose of this thesis, the results are categorised into two parts: those derived from the QAE and those from the VQC. The following subsections elaborate on these results with detailed analyses of their implications for QML in HEP.

7.8.1. Anomaly Detection Performance of the QAE

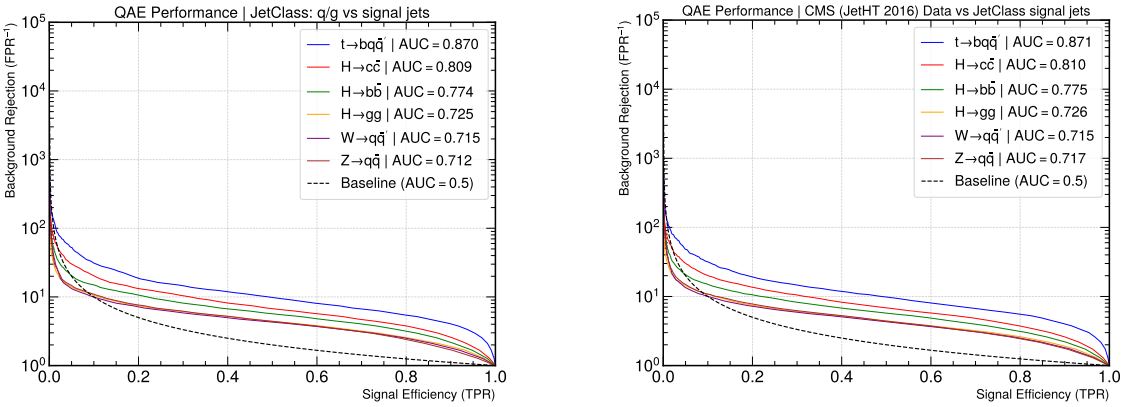
The QAE, as detailed in Section 7.7.3, is trained using both simulated background and real experimental data. In the absence of a suitable simulation dataset containing exotic BSM physics signatures, the generation of which exceeded the scope of this thesis, jets produced by established SM processes and already present in the JetClass dataset were employed as proxy signals for evaluating the QAE's anomaly detection capabilities. These processes include: $t \to bq\overline{q'}$, $\mathrm{H} \to b\bar{b}$, $\mathrm{H} \to c\bar{c}$, $\mathrm{H} \to gg$, $\mathrm{W} \to q\overline{q'}$ and $\mathrm{Z} \to q\overline{q'}$. The selection of these particular processes is justified by their characteristic multi-prong signatures, which exhibit substructure patterns distinct from the predominantly one-prong substructure observed in jets initiated by light quarks or gluons (constituting the background sample). This inherent topological difference provides an appropriate testing ground for assessing the discriminative power of the QAE.

These evaluations yield several significant observations:

- The QAE demonstrates strong performance as an anomaly detection tool, exhibiting varying degrees of discriminatory power across all investigated signal categories against the background comprised of light quark- or gluon-initiated jets. This performance variation correlates with the distinctive substructure characteristics of each signal type, with the highest performance being observed for the three-pronged top quark decay. Figure 7.7 presents a comparative analysis of the performance metrics for QAEs trained on background simulation and experimental data, quantified through the AUC metric. The observed separation capability confirms the QAE's potential for identifying anomalous jet patterns with complex multi-pronged substructure.
- The results presented here are furthermore, to the best of the authors' knowledge, the first application of a QNN architecture to experimental data recorded by the CMS Detector. Notably, the QAE maintains consistent performance levels when tasked with distinguishing between simulated signal jets and experimental data predominantly composed of light quark or gluon jets. This consistency substantiates the robustness of the 1P1Q encoding methodology against potential biases introduced through detector simulation processes, a critical consideration for practical applications in experimental settings.
- Furthermore, the QAE trained exclusively on experimental data exhibits negligible separation capability between real data and simulated background samples, as evidenced by the loss distributions shown in Figure 7.8. This observation corroborates the unbiased nature of the quantum encoding, suggesting that the representation learned by the QAE correspond to genuine physical characteristics rather than simulation artefacts. Such performance consistency across simulated and experimental domains provide a clearer pathway to deployment in real-world anomaly detection scenarios.

These findings collectively demonstrate the viability of quantum machine learning algorithms, specifically using the 1P1Q encoding and the algorithms described in this thesis, for anomaly detection in HEP. These results suggest potential advantages for identifying unknown BSM phenomena in future analyses of CMS data.

The QAE was trained for several choices of input qubits and latent space dimension. The improvements when increasing the number of particles beyond 6 are found to be marginal, but non-zero, which is to be expected since a large part of the information is expected to be carried by the hardest jet constituents. However, a larger number of jet constituents are expected to



(a) Performance of the QAE trained on simulated MC background **(b)** Performance of the QAE trained on the CMS JetHT (2016) Dataset

Figure 7.7.: Performance of a QAE with 10 input qubits and a latent space of 2 qubits, trained on simulated background (left) and real data (right).

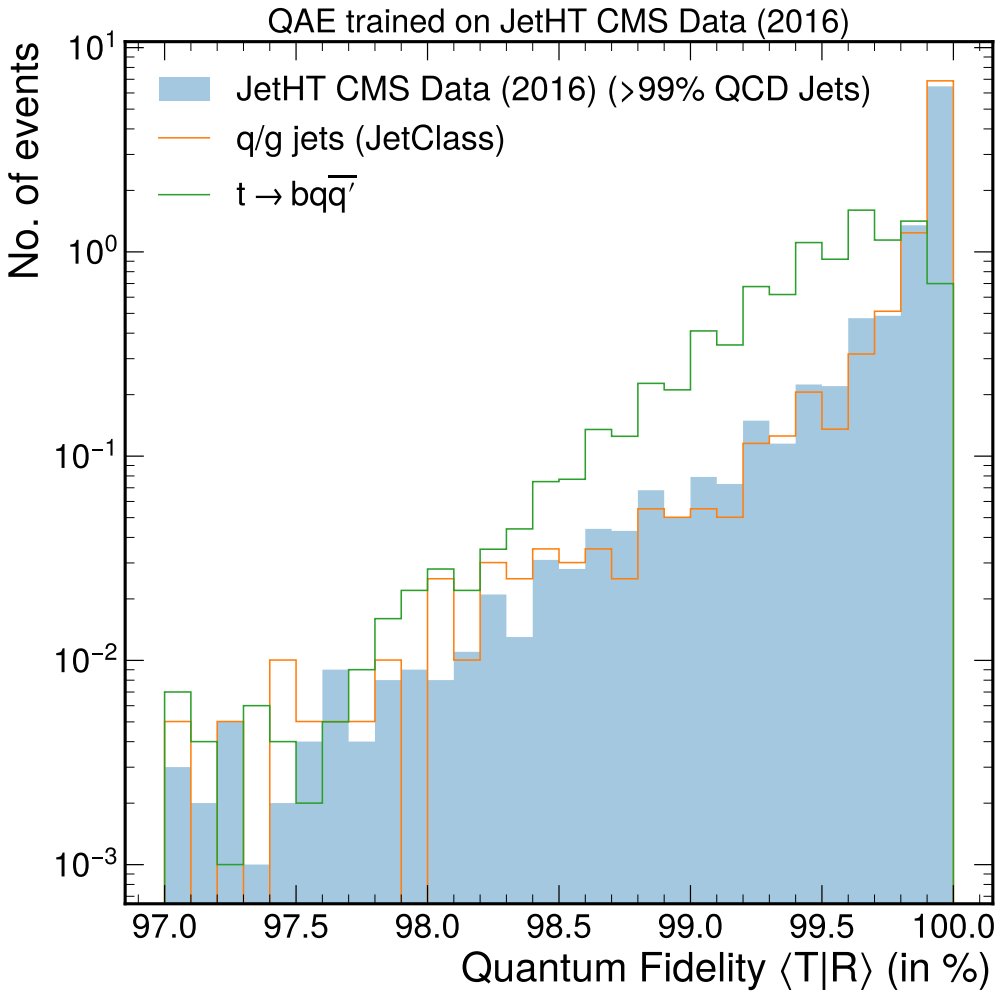


Figure 7.8.: Fidelity Distributions for a QAE trained on real data from the CMS Detector, when applied to data, simulated background and simulated jets arising from a top-quark decay

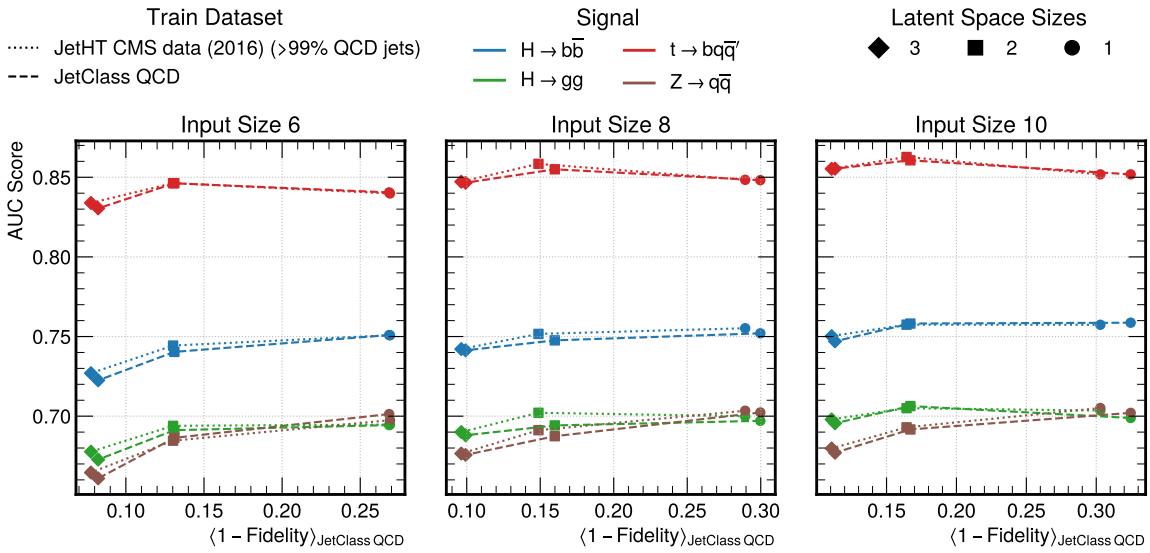


Figure 7.9.: AUC vs (1-Fidelity) (see Equation 7.11) for QAEs trained on 6,8 and 10 input particles respectively, on both real data (dotted) and simulated background (dashed). The middle ground is found to exist at a latent space dimension of 2

allow the algorithm to model the jet substructure more accurately, which explains the slight increase in performance. The best performance for all choices of input particle number, was found to be with a latent space corresponding to 2 qubits, which strikes an adequate balance between too much information (corresponding to higher reconstruction fidelity values) and too little (corresponding to lower reconstruction fidelity values). For $N = 6, 8, 10$ qubits, these results in terms of the metric AUC are presented in Figure 7.9. The best performing model is therefore the one with 10 input qubits and a latent space dimension of 2, which is therefore the one for which all other results are shown in this thesis.

7.8.1.1. Classical Benchmark for the QAE

The QAE employed in the 1P1Q approach is benchmarked against a classical autoencoder architecture, as described in Section 4.6. This classical implementation comprises two distinct DNNs functioning as the encoder and decoder respectively. These networks perform sequentially a dimensionality reduction and subsequently a reconstruction of the input jet constituent kinematics.

For a configuration with N input jet constituents (corresponding to N input qubits in the QAE), the classical autoencoder requires $3N$ input nodes, representing the three-dimensional kinematic vectors (p_T, η, ϕ) of each jet constituent. To maintain methodological equivalence and ensure a fair comparison between architectures, the classical autoencoder's latent space incorporates 3 nodes for each qubit in the QAE's latent representation.

The encoder component of the classical architecture implements a DNN with three hidden layers that progressively reduce dimensionality through a 20-16-12 node configuration, leading up to the latent space representation. Correspondingly, the decoder performs the inverse transformation through three hidden layers with a 12-16-20 node structure, ultimately restoring the original input dimensionality. Training of the classical autoencoder proceeds through minimisation of the Mean Squared Error (MSE) between input and output jet representations,

both expressed as flattened arrays of shape $(3N,)$. Identical feature sets are utilised for both the quantum and classical training procedures.

Performance evaluation, quantified through the AUC metric, is summarised in Table 7.1 for multiple benchmark signal processes. These results demonstrate that the QAE, despite possessing only 31 trainable parameters, consistently outperforms its classical counterpart, which incorporates approximately $O(3000)$ trainable parameters.

Table 7.1.: Comparison of anomaly detection performance (AUC) between the QAE and the Classical Autoencoder (CAE) across various signal processes.

Model	Signals					
	$Z \to q\bar{q}$	$W \to q\bar{q}$	$H \to gg$	$H \to b\bar{b}$	$H \to c\bar{c}$	$t \to bq\bar{q}$
QAE	0.715	0.715	0.729	0.774	0.810	0.872
CAE	0.676	0.675	0.740	0.739	0.767	0.858

This performance disparity underscores a fundamental advantage of quantum approaches in this domain. The QAE achieves superior discrimination capabilities with approximately two orders of magnitude fewer parameters than its classical counterpart (32 vs $O(3000)$), suggesting significantly enhanced parameter efficiency. This remarkable result derives from the quantum architecture's inherent ability to exploit high-dimensional Hilbert spaces and quantum phenomena such as superposition and entanglement to capture complex data patterns. Beyond performance metrics alone, the QAE's reduced parameterisation offers substantial advantages in terms of model interpretability and visualisation. With fewer parameters, the quantum operations can be more readily visualised and understood, facilitating deeper understanding of the learning process. This interpretability advantage contrasts with classical DNNs, which often function as effective but opaque "black boxes" due to the large number of parameters contained therein. Additionally, the parameter efficiency demonstrated here holds particular promise for deployment in environments where computational efficiency is of great importance. Ultimately, approaches using QML may offer significant practical advantages for anomaly detection in HEP beyond simple performance improvements.

7.8.2. Supervised Classification Performance of the VQC

The VQC implemented in this work functions as a specialised jet classifier, colloquially termed a *top tagger*, which is trained to discriminate between jets originating from top quark decays (characterised by their distinctive three-pronged substructure) and those initiated by light quarks or gluons (typically exhibiting a simple one-pronged topological structure). While Section 7.7.3 presents a comprehensive description of the training methodology and datasets employed, this section focuses exclusively on the experimental results and their implications for the usage of QML in jet classification.

For rigorous evaluation of the VQC's efficacy, a comparison is conducted against the current state-of-the-art classical jet classification model, ParticleTransformer [82]. Although ParticleTransformer was originally designed as a multi-class classifier capable of simultaneously differentiating up to 10 distinct jet categories, it is constrained to function as a binary classifier in this comparative analysis to maintain parity with the VQC. Furthermore, to ensure a fair comparison based on equivalent input information, the ParticleTransformer is restricted to utilise only the four-momentum components (p_x, p_y, p_z, E) of jet constituents, although its

conventional implementation makes use of several additional features such as charge and particle IDs. Both models are trained on an identical and computationally less intensive dataset of 1000 jets, ensuring comparability of results.

The performance evaluation employs two complementary metrics. First, the AUC provides an assessment of classification performance across the range of threshold values. Second, the background rejection factor at a specified signal efficiency, denoted as REJ_X [82], is calculated according to:

$$\mathrm{REJ}_X = \frac{1}{\mathrm{FPR}}, \qquad \text{at} \quad \mathrm{TPR} = X\%, \tag{7.33}$$

where FPR represents the False Positive Rate and TPR denotes the True Positive Rate. This metric quantifies the classifier's ability to reject background jets while maintaining a specified level of signal efficiency, a crucial performance indicator for practical HEP applications where obtaining high signal purity is often important.

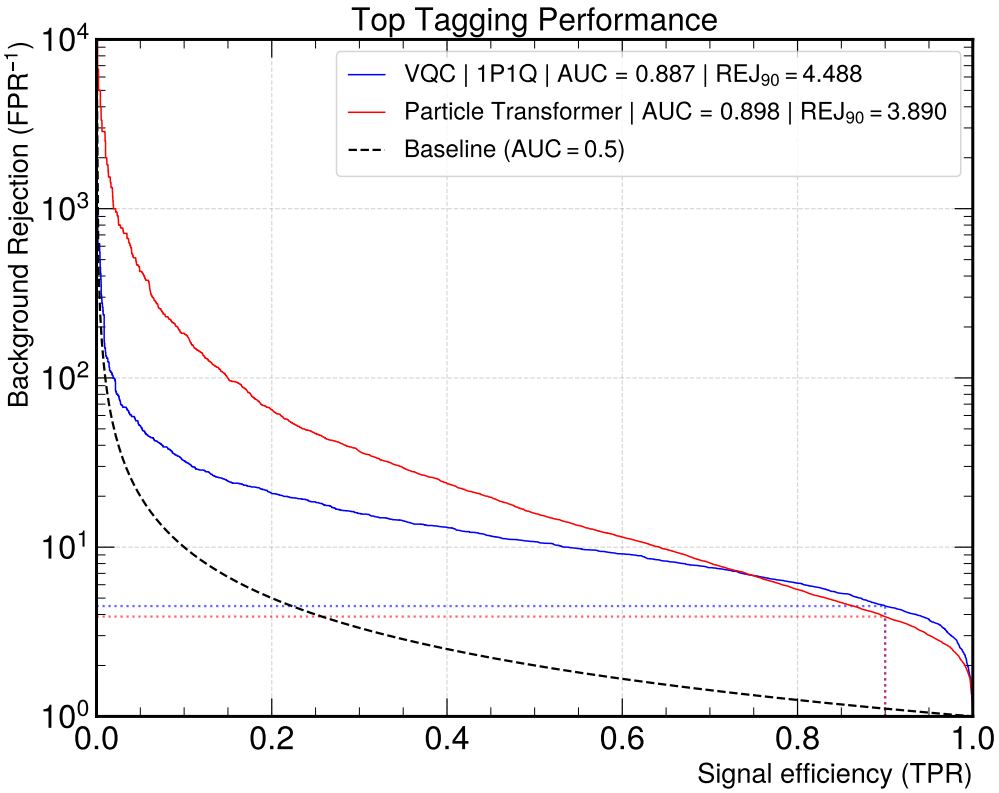


Figure 7.10.: Receiver Operating Characteristic (ROC) curves comparing the performance of the VQC and ParticleTransformer for top quark jet identification. The comparable performance is achieved despite the VQC having approximately five orders of magnitude fewer parameters. Also highlighted is the high signal efficiency region where the VQC demonstrates superior background rejection.

The results yield several insights regarding the 1P1Q approach when used for supervised classification:

- The VQC demonstrates robust performance as a top-tagger, achieving an AUC of 0.887. This performance metric exceeds that attained by the QAE for top quark decay jet identification, a predictable outcome considering the VQC's supervised learning paradigm, which

benefits from seeing both background and signal samples during training, enhancing its classification capabilities relative to unsupervised algorithms.

- Perhaps more remarkable is the VQC's competitive performance relative to the ParticleTransformer, as illustrated in Figure 7.10. This competitive capability is achieved despite a stark disparity in model complexity—the VQC incorporates merely 32 trainable parameters, in contrast to the transformer-based architecture's extensive parameterisation of $O(10^6)$ parameters. This five-order-of-magnitude reduction in parameter count represents a transformative advantage for model interpretability, computational efficiency, and potential hardware implementation.
- The VQC's performance advantage becomes particularly pronounced in the high signal efficiency regime, where it marginally surpasses the ParticleTransformer. Specifically, at a signal efficiency of 90% (a practical choice of the working point), the VQC achieves a background rejection factor (REJ_{90}) of 4.488, exceeding the ParticleTransformer's value of 3.890. This performance in the high efficiency regime suggests that the quantum approach may potentially offer advantages in scenarios where maintaining high signal purity is important.

These findings, from both the QAE and the VQC underscore the considerable potential of QML in HEP. The VQC's ability to achieve, and in specific regimes, even exceed the performance of a highly sophisticated classical model while utilising far fewer parameters suggests a fundamental advantage in the quantum approach to feature extraction and subsequent classification. This efficiency furthermore translates to enhanced interpretability, and potentially mitigates overfitting risks, enabling effective training with smaller datasets.

7.9. Closing Studies: Feature Ablation

The efficacy of QML architectures is significantly influenced by both the data encoding strategy and the quantum circuit architecture. To rigorously evaluate the 1P1Q encoding methodology employed in the QAE and VQC implementations, a systematic feature ablation analysis was conducted. This investigation quantifies the contribution of individual kinematic features to the overall performance, and demonstrates the optimal usage of all features by the QNNs.

The ablation procedure proceeds by excluding specific particle features from the encoding scheme and observing the resultant impact on performance. When eliminating an angular variable such as η or ϕ from the encoding, the corresponding rotation gate is entirely removed from the quantum circuit (see Equation 7.25). In the case of the p_T ablation, the scaling factor f (see Equations 7.22 and 7.23) is replaced with a constant value of 1.

Table 7.2.: Performance comparison (AUC scores) for different input feature combinations in the QAE and VQC models. The benchmark signal is $t \to bq\bar{q}$. Both models were trained using 10,000 simulated jets from the JetClass dataset.

Model	Inputs			
	(p_T, η, ϕ)	(p_T, η)	(η, ϕ)	(p_T, ϕ)
QAE	0.872	0.825	0.823	0.827
VQC	0.886	0.851	0.808	0.857

As illustrated in Table 7.2, the omission of any individual kinematic feature results in a measurable degradation of classification performance for both quantum models. This degradation exhibits model-specific patterns that provide insight into the feature utilisation mechanisms of the respective architectures. The most pronounced performance reduction is observed when p_T is excluded, more so in the case of the VQC. The ROC curves corresponding to these ablation studies for the VQC are presented in Figure 7.11. Interestingly, this performance degradation

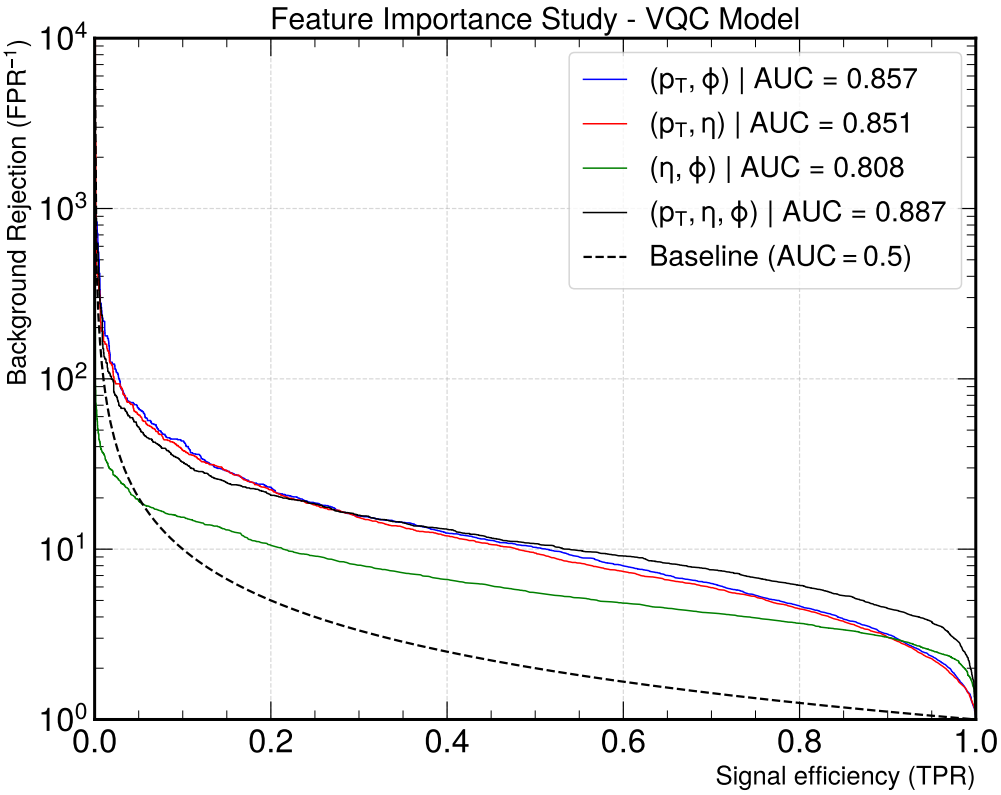


Figure 7.11.: ROC curves comparing the performance of the VQC for different sets of input features. A gradual drop in performance is observed when features are dropped, with the most significant reduction coming from the exclusion of p_T information.

appears less severe across features in the QAE compared to the VQC. This is hypothesised to stem from the QAE's inherent compression mechanism, which potentially facilitates a more balanced utilisation of all input features through the distribution of information across entangled quantum states.

These studies highlight the particular significance of p_T information in jet classification or anomaly detection tasks. Furthermore, they demonstrate that the 1P1Q encoding methodology effectively captures complementary information from both momentum and angular variables. The consistent pattern of performance degradation observed across both quantum models upon feature reduction substantiates the suitability and information efficiency of the 1P1Q encoding scheme for high-energy physics applications.

These findings suggest that optimal performance in quantum jet classification necessitates the complete triplet of kinematic features (p_T, η, ϕ). Therefore, a comprehensive encoding approach like 1P1Q ensures that the QNNs can utilise all available information for both supervised and unsupervised tasks.

7.10. Outlook for QML in HEP

The results presented in this thesis demonstrate that QML techniques offer promising avenues for future HEP analyses. It has been shown that an intuitive encoding strategy like 1P1Q, when coupled with quantum circuit-based algorithms, can produce powerful tools for both anomaly detection and supervised classification. These approaches achieve competitive or superior performance compared to classical counterparts while utilising significantly fewer parameters, thus enhancing interpretability and computational efficiency. The inherent ability of quantum circuits to exploit the geometric nature of input data is particularly advantageous for jet physics applications, where the spatial and kinematic relationships between particles contain important information. Quantum operations naturally preserve these relationships through unitary transformations, enabling effective feature extraction and pattern recognition in high-dimensional data.

Despite these encouraging results, substantial work remains to be undertaken regarding the fundamental understanding of the learning processes within parameterised quantum circuits. The interpretability of quantum operations and their relationship to physical observables requires further theoretical development. Additionally, the optimisation landscapes of variational quantum circuits require a rigorous characterisation to address challenges such as barren plateaus and local minima that can hinder the training process. As quantum hardware continues to evolve and error mitigation techniques improve, the practical implementation of QML algorithms in HEP will become more and more viable. The confluence of quantum computing and experimental HEP lays the ground for interdisciplinary research that may ultimately transform how physics analyses are conducted in the post-NISQ era.

The work presented here in this thesis is also available as a preprint [200] on arXiv, currently undergoing peer review for publication. Two master theses [210, 211] have also been written as part of this project.

Part V.

Conclusion and Appendices

8. Conclusion

This thesis presents two novel approaches for applying Machine Learning (ML) in High Energy Physics (HEP), demonstrating the potential of both classical and quantum algorithms for discovering new physics at the LHC. The first part introduces the Variational Autoencoder-Quantile Regression (VAE-QR) methodology, a model-agnostic approach for searching for resonant anomalies in dijet events recorded by the CMS detector at $\sqrt{s} = 13\,\text{TeV}$ during Run 2 of the LHC. The VAE is trained exclusively on jets from dijet events in a control region dominated by QCD multijet background events. Notably, this is a fully data-driven approach that does not require the usage of any MC simulation. When applied to events in the signal region, the reconstruction score serves as the basis for constructing an anomaly metric that is decorrelated from the dijet invariant mass (m_{jj}) using a DNN-based QR approach. A comprehensive search for resonant anomalies performed over the m_{jj} range from 1.8 to 6 TeVreveals no significant deviations from the Standard Model (SM) expectation, with the highest observed local significance being 2.3σ at 4.9 TeV. Subsequently, exclusion limits are set on a diverse range of signal models of the form $A \rightarrow BC$, where a heavy resonance A decays to two lighter particles B and C with complex topologies ranging from 1 + 2 prong to 6 + 6 prong structures. Almost all of these benchmark signals on which exclusion limits are set represent the first such limits obtained at the CMS Experiment, with these groundbreaking results now published in the journal *Reports on Progress in Physics* [51]. The limits obtained and the expected sensitivities to a broad range of signal models are shown to be substantially better than those from the inclusive search, where no anomaly detection cut is applied. Furthermore, all methods part of this analysis, including the VAE-QR approach presented in this thesis, demonstrate superior performance compared to traditional substructure variable-based cuts, establishing it as a powerful new tool for Beyond the Standard Model (BSM) physics searches. The intrinsically model-agnostic and data-driven nature of the VAE-QR method enables its application to a far broader class of signal models than demonstrated here, with preliminary studies indicating sensitivity to broad-width resonance signals as well. Figure 8.1 for example, shows the most anomalous event, in terms of the VAE-QR anomaly score, as recorded during the Run 2 data taking period of the CMS Detector.

The second part of this thesis introduces a Quantum Machine Learning (QML)-based approach to anomaly detection in HEP, with the 1P1Q encoding scheme that maps individual jet constituent information onto two-level quantum states (qubits) through quantum unitary operations. The powerful performance of this encoding is demonstrated through two quantum algorithms: the Quantum Autoencoder (QAE) for anomaly detection and the Variational Quantum Classifier (VQC) for binary classification. The QAE, trained on jets initiated by light quarks or gluons with single-pronged topology, demonstrates superior performance compared to a benchmark classical autoencoder across a wide range of signal models, despite utilising only $\mathcal{O}(1\%)$ of the trainable parameters. Specifically, with ten jet constituents and a training sample of ten thousand jets, the QAE achieves an Area Under the Curve of 0.872 for distinguishing top quark-initiated jets from light quark- or gluon-initiated jets, compared to 0.858 for the classical autoencoder. The VQC demonstrates competitive performance for binary classification

between top quark-initiated and light quark- or gluon-initiated jets, achieving an AUC of 0.886 with merely thirty-two trainable parameters, each such parameter corresponding to a rotation visualisable in three-dimensional space. In comparison, the state-of-the-art ParticleTransformer algorithm [82] against which it is benchmarked, contains $\mathcal{O}(10^6)$ trainable parameters. Notably, the VQC is able to exceed the performance of the ParticleTransformer in the high signal efficiency region ($> 80\%$), whilst offering substantially enhanced interpretability. The results presented in this thesis are currently available as a preprint [200] and are, as of June 2025, undergoing peer-review for publication in a journal.

These results underscore the potential of machine learning, both classical and quantum, for future analyses in HEP, particularly within LHC experiments such as the CMS Experiment. Unsupervised machine learning techniques serve as a powerful complementary approach to traditional dedicated searches for BSM physics phenomena. Unlike targeted searches that are optimised for specific theoretical models and signatures, these data-driven methods require minimal assumptions about the signal characteristics, enabling them to probe a substantially broader phase space. This model-agnostic nature allows unsupervised approaches to potentially identify any statistically significant deviation from SM predictions, regardless of whether such deviations correspond to theoretically motivated scenarios or represent entirely unexpected phenomena. While dedicated searches remain essential for testing specific theoretical predictions with optimal sensitivity, unsupervised methods cast a safety net against the possibility that new physics might manifest in ways other than those anticipated by current theoretical frameworks. The synergy between these complementary approaches, the first being targeted searches providing deep sensitivity to specific models and the second being unsupervised methods offering broad coverage across the entire accessible phase space, together serve to maximise the discovery potential for BSM physics at the LHC.

The inherent limitations of Noisy Intermediate-Scale Quantum devices present significant obstacles to the practical implementation of QML algorithms. These hardware constraints include low coherence times, errors in gate fidelity, and the presence of quantum noise, which collectively degrade the performance of quantum circuits. This necessitates the usage of complex error mitigation strategies to maintain adequate accuracy. The optimisation landscape of quantum NNs is not completely understood, with phenomena such as barren plateaus hampering the training convergence of QML algorithms. Additionally, the nature of current hardware and the overhead associated with error correction protocols limit the depth and complexity of QML models that can be feasibly implemented on near-term quantum devices. These challenges, combined with the high operational costs of quantum hardware, make the deployment of QML algorithms on current devices technically and economically prohibitive for most applications. At the same time, QML promises exponential speedup and dramatic reductions in parameter complexity, along with far greater interpretability compared to classical ML approaches, particularly in the post-NISQ era. This advantage could prove crucial as detector granularity and data rates continue to increase at future colliders, where the computational demands of classical algorithms may become prohibitive. The application of such advanced machine learning algorithms to the analysis of data arising from current and future particle collider experiments, with their unprecedented energies and luminosities, presents an exciting frontier for discovering new physics that may lie beyond the SM.

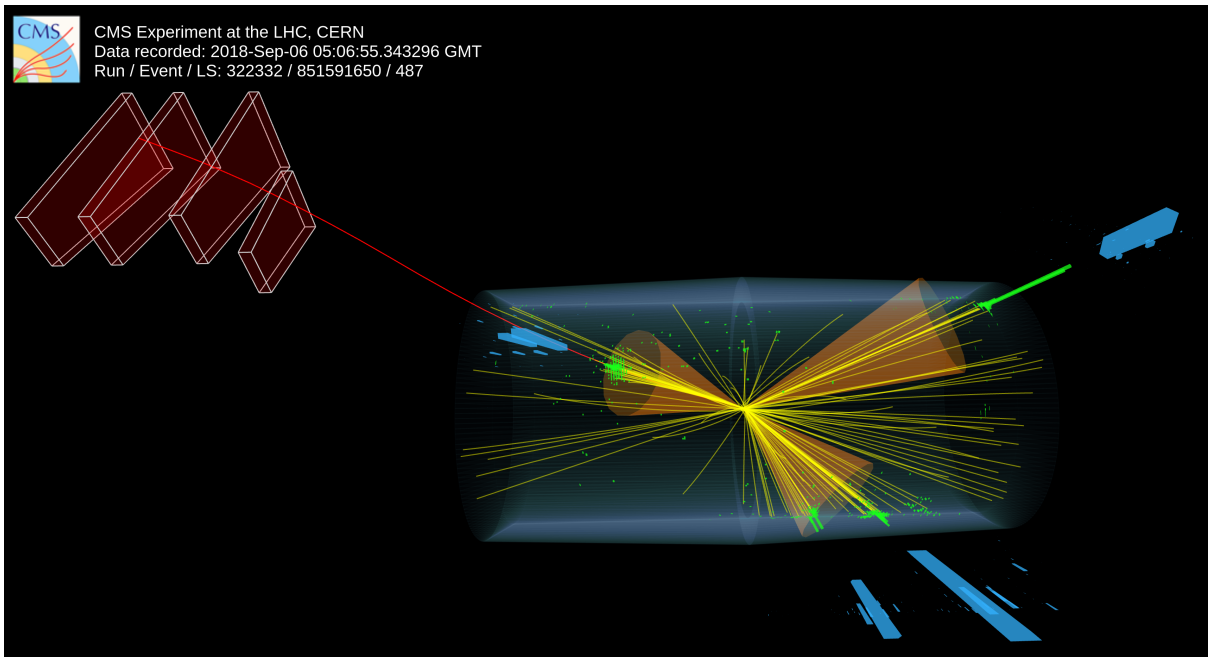


Figure 8.1.: The most anomalous event of Run 2 (2016-18) at the CMS Experiment, as identified by the VAE-QR algorithm presented in this thesis.

Bibliography

[1] S. L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nucl. Phys.* 22 (1961), p. 579. DOI: 10.1016/0029-5582(61)90469-2.

[2] S. Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), p. 1264. DOI: 10.1103/PhysRevLett.19.1264.

[3] A. Salam. "Weak and Electromagnetic Interactions". In: *Elementary Particle Theory*. Ed. by N. Svartholm. Stockholm: Almqvist & Wiksell, 1968.

[4] G. Aad and others (ATLAS Collaboration). "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), p. 1. DOI: 10.1016/j.physletb.2012.08.020.

[5] S. Chatrchyan and others (CMS Collaboration). "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Phys. Lett. B* 716 (2012), p. 30. DOI: 10.1016/j.physletb.2012.08.021.

[6] S. et al Chatrchyan. "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV". In: *Journal of High Energy Physics* 2013.6 (). ISSN: 1029-8479. DOI: 10.1007/jhep06(2013)081. URL: http://dx.doi.org/10.1007/JHEP06(2013)081.

[7] Andrew Blance and Michael Spannowsky. "Quantum Machine Learning for Particle Physics using a Variational Quantum Classifier". In: *JHEP* 02 (2021), p. 212. DOI: 10.1007/JHEP02(2021)212. arXiv: 2010.07335 [hep-ph].

[8] V. S. Ngairangbam, M. Spannowsky, and M. Takeuchi. "Anomaly detection in high-energy physics using a quantum autoencoder". In: *Physical Review D* 105.9 (2022), p. 095004. DOI: 10.1103/PhysRevD.105.095004.

[9] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[10] P. W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Phys. Lett.* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.

[11] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[12] H. Georgi and S. L. Glashow. "Unity of All Elementary Particle Forces". In: *Phys. Rev. Lett.* 32 (1974), p. 438. DOI: 10.1103/PhysRevLett.32.438.

[13] M. Gell-Mann. "A Schematic Model of Baryons and Mesons". In: *Phys. Lett.* 8 (1964), p. 214. DOI: 10.1016/S0031-9163(64)92001-3.

[14] G. Zweig. *An SU(3) Model for Strong Interaction Symmetry and Its Breaking*. Tech. rep. 8182/TH.401. CERN, 1964.

[15] G. 't Hooft. "Renormalizable Lagrangians for Massive Yang-Mills Fields". In: *Nucl. Phys. B* 35 (1971), pp. 167–188. DOI: 10.1016/0550-3213(71)90303-5.

[16] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory*. Reading, MA: Perseus Books, 1995.

[17] Y. Fukuda and others (Super-Kamiokande Collaboration). "Evidence for Oscillation of Atmospheric Neutrinos". In: *Phys. Rev. Lett.* 81 (1998), p. 1562. DOI: 10.1103/PhysRevLett.81.1562.

[18] Q. R. Ahmad and others (SNO Collaboration). "Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory". In: *Phys. Rev. Lett.* 89 (2002), p. 011301. DOI: 10.1103/PhysRevLett.89.011301.

[19] G. Arnison and others (UA1 Collaboration). "Experimental Observation of Lepton Pairs of Invariant Mass around 95 GeV/c^2 at the CERN SPS Collider". In: *Phys. Lett. B* 126 (1983), p. 398. DOI: 10.1016/0370-2693(83)90605-2.

[20] D. J. Gross and F. Wilczek. "Ultraviolet Behavior of Non-Abelian Gauge Theories". In: *Phys. Rev. Lett.* 30 (1973), p. 1343. DOI: 10.1103/PhysRevLett.30.1343.

[21] H. D. Politzer. "Reliable Perturbative Results for Strong Interactions?" In: *Phys. Rev. Lett.* 30 (1973), p. 1346. DOI: 10.1103/PhysRevLett.30.1346.

[22] P. A. Zyla and others (Particle Data Group). "Review of Particle Physics". In: *Prog. Theor. Exp. Phys.* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[23] F. Capozzi et al. "Status of three-neutrino oscillation parameters, circa 2013". In: *Phys. Rev. D* 89.9 (2014), p. 093018. DOI: 10.1103/PhysRevD.89.093018.

[24] I. Esteban et al. "On the determination of the solar neutrino mixing angle from globally available data". In: *J. High Energ. Phys.* 2020.9 (2020), p. 178. DOI: 10.1007/JHEP09(2020)178.

[25] G. Arnison and others (UA1 Collaboration). "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV". In: *Phys. Lett. B* 122 (1983), p. 103. DOI: 10.1016/0370-2693(83)91177-2.

[26] M. Banner and others (UA2 Collaboration). "Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN $\bar{p}p$ Collider". In: *Phys. Lett. B* 122 (1983), p. 476. DOI: 10.1016/0370-2693(83)91178-4.

[27] B. Aubert and others (BaBar Collaboration). "Observation of CP Violation in the B^0 Meson System". In: *Phys. Rev. Lett.* 87 (2001), p. 091801. DOI: 10.1103/PhysRevLett.87.091801.

[28] K. Abe and others (Belle Collaboration). "Observation of Large CP Violation in the Neutral B Meson System". In: *Phys. Rev. Lett.* 87 (2001), p. 091802. DOI: 10.1103/PhysRevLett.87.091802.

[29] Oliver Sim Brüning et al. *LHC Design Report*. Tech. rep. CERN-2004-003-V-1. CERN, 2004.

[30] Werner Herr and B Muratori. "Concept of luminosity". In: (2006). DOI: 10.5170/CERN-2006-002.361. URL: https://cds.cern.ch/record/941318.

[31] Bo Andersson et al. "Parton Fragmentation and String Dynamics". In: *Physics Reports* 97 (1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7.

[32] Bryan R. Webber. "A QCD Model for Jet Fragmentation Including Soft Gluon Interference". In: *Nuclear Physics B* 238 (1984), pp. 492–528. DOI: 10.1016/0550-3213(84)90355-3.

[33] Guido Altarelli and Giorgio Parisi. "Asymptotic Freedom in Parton Language". In: *Nuclear Physics B* 126 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4.

[34] Stefano Catani et al. "Longitudinally-Invariant k_T Clustering Algorithms for Hadron–Hadron Collisions". In: *Nuclear Physics B* 406 (1993), pp. 187–224. DOI: 10.1016/0550-3213(93)90166-M.

[35] Stephen D. Ellis and Davison E. Soper. "Successive Combination Jet Algorithm for Hadron Collisions". In: *Physical Review D* 48 (1993), pp. 3160–3166. DOI: 10.1103/PhysRevD.48.3160.

[36] Gavin P. Salam. "Towards Jetography". In: *The European Physical Journal C* 67 (2010), pp. 637–686. DOI: 10.1140/epjc/s10052-010-1314-6.

[37] Yuri L. Dokshitzer et al. "Better Jet Clustering Algorithms". In: *Journal of High Energy Physics* 08 (1997), p. 001. DOI: 10.1088/1126-6708/1997/08/001.

[38] Matthias Wobisch and Thorsten Wengler. "Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering". In: *arXiv preprint hep-ph/9907280* (1998). eprint: hep-ph/9907280.

[39] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_T Jet Clustering Algorithm". In: *Journal of High Energy Physics* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063.

[40] George Sterman and Steven Weinberg. "Jets from Quantum Chromodynamics". In: *Physical Review Letters* 39 (1977), pp. 1436–1439. DOI: 10.1103/PhysRevLett.39.1436.

[41] Yuri L. Dokshitzer and Steven I. Troyan. "Better Jet Clustering Algorithms". In: *Proceedings of the Workshop on Jet Studies at LEP and HERA* (1991). Published in J. Phys. G17:1602-1604, 1991.

[42] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual". In: *European Physical Journal C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2.

[43] Andrew J. Larkoski, Ian Moult, and Benjamin Nachman. "Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning". In: *Physics Reports* 841 (2020), pp. 1–63. DOI: 10.1016/j.physrep.2019.11.001.

[44] ATLAS Collaboration. *ATLAS Technical Design Report*. Tech. rep. CERN-LHCC-99-15. CERN, 1999.

[45] CMS Collaboration. *The CMS Experiment at the CERN LHC*. Tech. rep. CERN-LHCC-2008-004. CERN, 2008.

[46] ALICE Collaboration. *ALICE: Technical Design Report of the Time Projection Chamber*. Tech. rep. CERN-LHCC-95-71. CERN, 1995.

[47] LHCb Collaboration. *LHCb Technical Design Report*. Tech. rep. CERN-LHCC-2003-030. CERN, 2003.

[48] CMS Collaboration. *The Phase-1 Upgrade of the CMS Level-1 Trigger*. Tech. rep. CERN-LHCC-2016-019. CERN, 2016.

[49] ATLAS Collaboration. *ATLAS Level-1 Trigger: Phase-I Upgrade Technical Design Report*. Tech. rep. CERN-LHCC-2017-020. CERN, 2017.

[50] ATLAS and CMS Collaborations. "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV". In: *Phys. Rev. Lett.* 114 (2015), p. 191803. DOI: 10.1103/PhysRevLett.114.191803.

[51] CMS Collaboration. “Model-agnostic search for dijet resonances with anomalous jet substructure in proton-proton collisions at s = 13 TeV”. In: *Reports on Progress in Physics* (2025). URL: http://iopscience.iop.org/article/10.1088/1361-6633/add762.

[52] CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[53] CMS Collaboration. *The CMS Tracker System Project: Technical Design Report*. Tech. rep. CERN-LHCC-98-006, CMS-TDR-5. CERN, 1998. URL: https://cds.cern.ch/record/368412.

[54] CMS Collaboration. “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s}$ = 7 TeV”. In: *JINST* 7 (2012), P10002. DOI: 10.1088/1748-0221/7/10/P10002.

[55] CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *JINST* 9 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009.

[56] CMS Collaboration. *The CMS Electromagnetic Calorimeter Project: Technical Design Report*. Tech. rep. CERN-LHCC-97-033, CMS-TDR-4. CERN, 1997. URL: https://cds.cern.ch/record/349375.

[57] CMS Collaboration. *The CMS Hadron Calorimeter Project: Technical Design Report*. Tech. rep. CERN-LHCC-97-031, CMS-TDR-2. CERN, 1997. URL: https://cds.cern.ch/record/357153.

[58] CMS Collaboration. *The CMS Muon Project: Technical Design Report*. Tech. rep. CERN-LHCC-97-032, CMS-TDR-3. CERN, 1997. URL: https://cds.cern.ch/record/343814.

[59] The CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (Oct. 2014), P10009–P10009. ISSN: 1748-0221. DOI: 10.1088/1748-0221/9/10/p10009. URL: http://dx.doi.org/10.1088/1748-0221/9/10/P10009.

[60] CMS Collaboration. *The TriDAS Project: Technical Design Report, Volume 1 – The Level-1 Trigger*. Tech. rep. CERN-LHCC-2000-038, CMS-TDR-6.1. CERN, 2000. URL: https://cds.cern.ch/record/706847.

[61] CMS Collaboration. *The TriDAS Project: Technical Design Report, Volume 2 – Data Acquisition and High-Level Trigger*. Tech. rep. CERN-LHCC-2002-026, CMS-TDR-6. CERN, 2002. URL: https://cds.cern.ch/record/578006.

[62] R. Frühwirth, G. Adam, and A. Strandlie. “The CMS Particle Flow Algorithm”. In: *Comput. Phys. Commun.* 140 (2001), pp. 197–214. DOI: 10.1016/S0010-4655(01)00293-0.

[63] *Particle–Flow Event Reconstruction in CMS and Performance for Jets, Taus, and Missing MET*. CMS Physics Analysis Summary CMS-PAS-PFT-09-001. CERN, 2009. URL: https://cds.cern.ch/record/1194487.

[64] CMS Collaboration. “Particle–Flow Reconstruction and Global Event Description with the CMS Detector”. In: *J. Inst.* 12 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003.

[65] D. Bertolini et al. “Pileup Per Particle Identification”. In: *JHEP* 10 (2014), p. 059. DOI: 10.1007/JHEP10(2014)059.

[66] F. Mokhtar et al. “Progress towards an Improved Particle Flow Algorithm at CMS with Machine Learning”. In: (2023). arXiv: 2303.17657.

[67] M. Vanadia. "Tracking, Particle Flow and Muon Performance at CMS and ATLAS". In: *Proc. 12th Large Hadron Collider Physics Conf. (LHCP 2024)*. 2024.

[68] CMS Collaboration. *High-precision measurement of the W boson mass with the CMS experiment at the LHC*. 2024. arXiv: 2412.13872 [hep-ex]. URL: https://arxiv.org/abs/2412.13872.

[69] CMS Collaboration. "Measurement of the top-quark mass in $t\bar{t}$ events with lepton+jets final states in pp collisions at $\sqrt{s}$ = 7 TeV". In: *JHEP* 12 (2012), p. 105. DOI: 10.1007/JHEP12(2012)105.

[70] CMS Collaboration. "Search for supersymmetry in proton-proton collisions at 13 TeV in final states with jets and missing transverse momentum". In: *JHEP* 10 (2021), p. 206. DOI: 10.1007/JHEP10(2021)206.

[71] G. Apollinari et al. "High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1". In: *CERN Yellow Reports: Monographs* 4 (2017), pp. 1–516. DOI: 10.23731/CYRM-2017-004.

[72] CMS Collaboration. *The Phase-2 Upgrade of the CMS Experiment*. Tech. rep. CERN-LHCC-2017-009. Technical Design Report. CERN, 2017. URL: https://cds.cern.ch/record/2283187.

[73] ATLAS Collaboration. *Technical Design Report for the ATLAS Inner Tracker Strip Detector*. Tech. rep. CERN-LHCC-2017-005. CERN, 2017. URL: https://cds.cern.ch/record/2257755.

[74] CMS Collaboration. *The Phase-2 Upgrade of the CMS Tracker*. Tech. rep. CERN-LHCC-2017-009. CERN, 2017. URL: https://cds.cern.ch/record/2272264.

[75] CMS Collaboration. *The Phase-2 Upgrade of the CMS Endcap Calorimeter*. Tech. rep. CERN-LHCC-2017-023. CERN, 2017. URL: https://cds.cern.ch/record/2293646.

[76] Gabriella Pásztor. "The Phase-2 Upgrade of the CMS Detector". In: *PoS* LHCP2022 (2023), p. 045. DOI: 10.22323/1.422.0045. URL: https://cds.cern.ch/record/2880161.

[77] Frank Rosenblatt. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain". In: *Psychological Review* 65.6 (1958), pp. 386–408. DOI: 10.1037/h0042519.

[78] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning Representations by Back-Propagating Errors". In: *Nature* 323 (1986), pp. 533–536. DOI: 10.1038/323533a0.

[79] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Vol. 25. 2012, pp. 1097–1105.

[80] CMS Collaboration. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *Journal of Instrumentation* 13 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158.

[81] Huilin Qu and Loukas Gouskos. "ParticleNet: Jet Tagging via Particle Clouds". In: *Phys. Rev. D* 101.5 (2020), p. 056019. DOI: 10.1103/PhysRevD.101.056019. arXiv: 1902.08570 [hep-ph].

[82] Huilin Qu, Congqiao Li, and Sitian Qian. "Particle Transformer for Jet Tagging". In: (Feb. 2022). arXiv: 2202.03772 [hep-ph].

[83] Javier Duarte and Jean-Roch Vlimant. "Graph Neural Networks for Particle Tracking and Reconstruction". In: *Artificial Intelligence for High Energy Physics*. World Scientific, 2022, pp. 387–436. arXiv: 2012.01249.

[84] Thea Aarrestad et al. "Fast convolutional neural networks on FPGAs with hls4ml". In: *Machine Learning: Science and Technology* 2.4 (July 2021), p. 045015. ISSN: 2632-2153. DOI: 10.1088/2632-2153/ac0ea1. URL: http://dx.doi.org/10.1088/2632-2153/ac0ea1.

[85] CMS ECAL Collaboration. "Autoencoder-Based Anomaly Detection System for Online Data Quality Monitoring of the CMS Electromagnetic Calorimeter". In: *Computing and Software for Big Science* (2023). arXiv: 2309.10157.

[86] Daniel Guest, Kyle Cranmer, and Daniel Whiteson. "Deep Learning and Its Application to LHC Physics". In: *Annual Review of Nuclear and Particle Science* 68 (2018), pp. 161–181. DOI: 10.1146/annurev-nucl-101917-021019. arXiv: 1806.11484.

[87] Alexander Radovic, Mike Williams, David Rousseau, et al. "Machine Learning at the Energy and Intensity Frontiers of Particle Physics". In: *Nature* 560 (2018), pp. 41–48. DOI: 10.1038/s41586-018-0361-2.

[88] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, et al. "Machine Learning and the Physical Sciences". In: *Reviews of Modern Physics* 91 (2019), p. 045002. DOI: 10.1103/RevModPhys.91.045002.

[89] CMS Collaboration. "End-to-end jet classification of quarks and gluons with the CMS Open Data". In: *Nucl. Instrum. Meth. A* 989 (2021), p. 164802. DOI: 10.1016/j.nima.2020.164802.

[90] Vasilis Belis, Patrick Odagiu, and Maurizio Pierini. "Machine Learning for Anomaly Detection in Particle Physics". In: *arXiv preprint arXiv:2312.14190* (2023).

[91] ATLAS Collaboration. "Dijet resonance search with weak supervision using 13 TeV pp collisions in the ATLAS detector". In: *Phys. Rev. Lett.* 125.13 (2020), p. 131801. DOI: 10.1103/PhysRevLett.125.131801.

[92] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *arXiv preprint* (2014). eprint: 1412.6980.

[93] Herbert Robbins and Sutton Monro. "A Stochastic Approximation Method". In: *Annals of Mathematical Statistics* 22.3 (1951), pp. 400–407. DOI: 10.1214/aoms/1177729586.

[94] David Guest, Kyle Cranmer, and Daniel Whiteson. "Deep Learning and its Application to LHC Physics". In: *Ann. Rev. Nucl. Part. Sci.* 68 (2018), pp. 161–181. DOI: 10.1146/annurev-nucl-101917-021019. arXiv: 1806.11484 [hep-ex].

[95] Thea K. Aarrestad et al. "The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics". In: *Rept. Prog. Phys.* 85.9 (July 2022), p. 094201. DOI: 10.1088/1361-6633/ac7623. arXiv: 2101.08320 [hep-ph].

[96] CMS Collaboration. "Search for new physics in multijet events using a deep autoencoder and significance-based anomaly detection". In: *Phys. Rev. D* 108.1 (2023), p. 012010. DOI: 10.1103/PhysRevD.108.012010. arXiv: 2212.00043 [hep-ex].

[97] Marco Farina, Yuichiro Nakai, and David Shih. "Searching for new physics with deep autoencoders". In: *Phys. Rev. D* 101.7 (2020), p. 075021. DOI: 10.1103/PhysRevD.101.075021. arXiv: 1808.08992 [hep-ph].

[98] Varun Chandola, Arindam Banerjee, and Vipin Kumar. "Anomaly Detection: A Survey". In: *ACM Comput. Surv.* 41.3 (2009), 15:1–15:58. DOI: 10.1145/1541880.1541882.

[99] Geoffrey E. Hinton and Ruslan R. Salakhutdinov. "Reducing the Dimensionality of Data with Neural Networks". In: *Science* 313.5786 (2006), pp. 504–507. DOI: 10.1126/science.1127647.

[100] Karl Pearson. "LIII. On Lines and Planes of Closest Fit to Systems of Points in Space". In: *Philosophical Magazine, Series 6* 2.11 (1901), pp. 559–572. DOI: 10.1080/14786440109462720.

[101] Hervé Bourlard and Yves Kamp. "Auto-association by multilayer perceptrons and singular value decomposition". In: *Biological Cybernetics* 59 (1988), pp. 291–294. DOI: 10.1007/BF00332918.

[102] Diederik P. Kingma and Max Welling. "Auto-Encoding Variational Bayes". In: *arXiv preprint* (2013). arXiv: 1312.6114 [stat.ML].

[103] Danilo J. Rezende, Shakir Mohamed, and Daan Wierstra. "Stochastic Backpropagation and Approximate Inference in Deep Generative Models". In: (2014), pp. 1278–1286. arXiv: 1401.4082 [stat.ML].

[104] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. New York: Wiley, 1991. ISBN: 978-0471062592.

[105] Irina Higgins et al. "β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework". In: *5th International Conference on Learning Representations (ICLR)*. Workshop track. 2017. URL: https://openreview.net/forum?id=Sy2fzU9gl.

[106] Olmo Cerri et al. "Variational Autoencoders for New Physics Mining at the Large Hadron Collider". In: *Journal of High Energy Physics* 2019.05 (2019), p. 036. DOI: 10.1007/JHEP05(2019)036. arXiv: 1811.10276 [hep-ex].

[107] Taoli Cheng et al. "Variational Autoencoders for Anomalous Jet Tagging". In: *Phys. Rev. D* 107 (2023), p. 016002. DOI: 10.1103/PhysRevD.107.016002.

[108] Pratik Jawahar et al. "Improving Variational Autoencoders for New Physics Detection at the LHC with Normalizing Flows". In: *Frontiers in Big Data* 5 (2022), p. 803685. DOI: 10.3389/fdata.2022.803685. arXiv: 2110.08508.

[109] Ian Goodfellow et al. "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems*. Vol. 27. 2014.

[110] Jared Barron et al. "Unsupervised Hadronic SUEP at the LHC". In: *JHEP* 12 (2021), p. 129. DOI: 10.1007/JHEP12(2021)129.

[111] Abhijith Gandrakota et al. "Real-time Anomaly Detection at the L1 Trigger of the CMS Experiment". In: *arXiv e-prints* (2024). arXiv: 2411.19506.

[112] *Model-agnostic search for dijet resonances with anomalous jet substructure in proton-proton collisions at $\sqrt{s}$ = 13 TeV*. Tech. rep. Geneva: CERN, 2024. URL: https://cds.cern.ch/record/2892677.

[113] Kinga Anna Wozniak. "Machine learning applications for new physics searches at CERN's Large Hadron Collider". PhD thesis. Vienna U., 2023. DOI: 10.25365/thesis.74519.

[114] Manuel Sommerhalder. "Developing and Applying Machine Learning Techniques for Model-Agnostic Searches for New Physics at the LHC". PhD thesis. 2024.

[115] Samuel Bright-Thonney. "Dark Photons & Deep Learning: Unconventional Searches for New Physics at the Large Hadron Collider". PhD thesis. Cornell U., 2024. DOI: 10.7298/28dx-dq85.

[116] CMS Collaboration. *CMS Luminosity – Public Results*. Tech. rep. CERN, 2021. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[117] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2014).

[118] Torbjörn Sjöstrand et al. *PYTHIA 8.3 Online Manual*. Accessed: 2025-05-12. 2024. URL: https://pythia.org/manuals/pythia8314/Welcome.html.

[119] Christian Bierlich et al. "A comprehensive guide to the physics and usage of PYTHIA 8.3". In: *SciPost Phys. Codebases* 8 (2022), 8–r8.3. arXiv: 2203.11601 [hep-ph].

[120] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *Journal of High Energy Physics* 2014.7 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep07(2014)079. URL: http://dx.doi.org/10.1007/JHEP07(2014)079.

[121] Paolo Nason. "A New method for combining NLO QCD with shower Monte Carlo algorithms". In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146 [hep-ph].

[122] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method". In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[123] Richard D. Ball et al. *An open-source machine learning framework for global analyses of parton distributions*. 2021. arXiv: 2109.02671 [hep-ph]. URL: https://arxiv.org/abs/2109.02671.

[124] A. M. et al Sirunyan. "Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements". In: *The European Physical Journal C* 80.1 (). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7499-4. URL: http://dx.doi.org/10.1140/epjc/s10052-019-7499-4.

[125] S. Agostinelli et al. "GEANT4 - A Simulation Toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[126] Ulrich Baur, Ian Hinchliffe, and Dieter Zeppenfeld. "Excited quark production at hadron colliders". In: *Int. J. Mod. Phys. A* 02 (1987), pp. 1285–1297. DOI: 10.1142/S0217751X87000661.

[127] Ulrich Baur, Michael Spira, and Peter Matthias Zerwas. "Excited-quark and -lepton production at hadron colliders". In: *Phys. Rev. D* 42 (1990), pp. 815–824. DOI: 10.1103/PhysRevD.42.815.

[128] Daniele Barducci et al. "Exploring Drell-Yan signals from the 4D Composite Higgs Model at the LHC". In: *J. High Energ. Phys.* 04 (2013), p. 152. DOI: 10.1007/jhep04(2013)152.

[129] Kaustubh Agashe et al. "Dedicated strategies for triboson signals from cascade decays of vector resonances". In: *Phys. Rev. D* 99 (2019), p. 075016. DOI: 10.1103/physrevd.99.075016.

[130] Yasuhiro Okada and Luca Panizzi. *LHC signatures of vector-like quarks*. 2012. arXiv: 1207.5607 [hep-ph]. URL: https://arxiv.org/abs/1207.5607.

[131] Mathieu Buchkremer et al. "Model-independent framework for searches of top partners". In: *Nuclear Physics B* 876.2 (Nov. 2013), pp. 376–417. ISSN: 0550-3213. DOI: 10.1016/j.nuclphysb.2013.08.010. URL: http://dx.doi.org/10.1016/j.nuclphysb.2013.08.010.

[132] Alexandra Carvalho. *Gravity particles from Warped Extra Dimensions, predictions for LHC*. 2018. arXiv: 1404.0102 [hep-ph].

[133] CMS Collaboration. "Search for a top or bottom quark decaying to a vector-like quark and a top or bottom quark in the all-jets final state at $\sqrt{s}$ = 13 TeV". In: *J. High Energ. Phys.* 09 (2022), p. 088. DOI: 10.1007/jhep09(2022)088.

[134] CMS Collaboration. "Search for Resonances Decaying to Three W Bosons in Proton-Proton Collisions at $\sqrt{s}$ = 13 TeV". In: *Phys. Rev. Lett.* 129 (2022), p. 021802. DOI: 10.1103/physrevlett.129.021802.

[135] CMS Collaboration. "Search for resonances decaying to three W bosons in the hadronic final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV". In: *Phys. Rev. D* 106 (2022), p. 012002. DOI: 10.1103/PhysRevD.106.012002.

[136] Y. Lecun et al. "Handwritten Digit Recognition with a Back-Propagation Network". In: *Advances in Neural Information Processing Systems 2*. Ed. by D. S. Touretzky. Morgan Kaufmann, 1990, pp. 396–404.

[137] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep Learning". In: *Nature* 521 (2015), pp. 436–444. DOI: 10.1038/nature14539.

[138] Haoqiang Fan, Hao Su, and Leonidas J. Guibas. "A Point Set Generation Network for 3D Object Reconstruction from a Single Image". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 2463–2471. DOI: 10.1109/CVPR.2017.264. arXiv: 1612.00603 [cs.CV].

[139] Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[140] Python Software Foundation. *Python Language Reference, version 3.10*. https://www.python.org. Accessed: 2025-06-02. 2021.

[141] Kaiming He et al. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification". In: *2015 IEEE International Conference on Computer Vision (ICCV)* (2015), pp. 1026–1034. DOI: 10.1109/ICCV.2015.123. arXiv: 1502.01852 [cs.CV].

[142] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)". In: *arXiv preprint arXiv:1511.07289* (2015).

[143] Jon Louis Bentley. "Multidimensional Binary Search Trees Used for Associative Searching". In: *Communications of the ACM* 18.9 (1975), pp. 509–517. DOI: 10.1145/361002.361007.

[144] James Dolen et al. "Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure". In: *Journal of High Energy Physics* 2016.05 (2016), p. 156. DOI: 10.1007/JHEP05(2016)156. arXiv: 1603.00027 [hep-ph].

[145] Andrew Blance, Michael Spannowsky, and Philip Waite. "Adversarially-trained autoencoders for robust unsupervised new physics searches". In: *Journal of High Energy Physics* 2019.10 (Oct. 2019). ISSN: 1029-8479. DOI: 10.1007/jhep10(2019)047. URL: http://dx.doi.org/10.1007/JHEP10(2019)047.

[146] Roger Koenker and Gilbert Jr. Bassett. "Regression Quantiles". In: *Econometrica* 46.1 (1978), pp. 33–50. DOI: 10.2307/1913643.

[147] James W. Taylor. "A Quantile Regression Neural Network Approach to Estimating the Conditional Density of Multiperiod Returns". In: *Journal of Forecasting* 19.4 (2000), pp. 299–311. DOI: 10.1002/1099-131X(200007)19:4<299::AID-FOR775>3.0.CO;2-V.

[148] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *European Physical Journal C* 71 (2011), p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0.

[149] S. S. Wilks. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". In: *Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62. DOI: 10.1214/aoms/1177732360. URL: https://projecteuclid.org/euclid.aoms/1177732360.

[150] S. Navas and others (Particle Data Group). "Review of Particle Physics". In: *Phys. Rev. D* 110.3 (2024). See Chapter 40: Statistics, revised August 2023 by G. Cowan, p. 030001. DOI: 10.1103/PhysRevD.110.030001. URL: https://pdg.lbl.gov/2024/reviews/rpp2024-rev-statistics.pdf.

[151] Kyle Cranmer. "Practical Statistics for the LHC". In: *arXiv preprint* (2015). arXiv: 1503.07622 [physics.data-an].

[152] J. S. Conway. "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra". In: *Proceedings of PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*. arXiv:1103.0354. 2011. URL: https://arxiv.org/abs/1103.0354.

[153] Wouter Verkerke and David Kirkby. "The RooFit toolkit for data modeling". In: *arXiv Physics e-prints* (2003). arXiv: physics/0306116 [physics.data-an].

[154] Alexander L. Read. "Presentation of search results: The CL_s technique". In: *Journal of Physics G: Nuclear and Particle Physics* 28 (2002), pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313.

[155] Thomas Junk. "Confidence level computation for combining searches with small statistics". In: *Nucl. Instrum. Meth. A* 434 (1999), pp. 435–443. DOI: 10.1016/S0168-9002(99)00498-2. arXiv: hep-ex/9902006 [hep-ex].

[156] Khachatryan et al. "Search for massive resonances in dijet systems containing jets tagged as W or Z boson decays in pp collisions at s $\sqrt{s}$ = 8 TeV". In: *Journal of High Energy Physics* 8 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep08(2014)173. URL: http://dx.doi.org/10.1007/JHEP08(2014)173.

[157] Ronald A. Fisher. "On the interpretation of χ^2 from contingency tables, and the calculation of P". In: *Journal of the Royal Statistical Society* 85.1 (1922), pp. 87–94. DOI: 10.2307/2340521.

[158] Mark J. Oreglia. "A Study of the Reactions $\psi' \to \gamma\gamma\psi$". SLAC Report SLAC-R-236. PhD thesis. Stanford University, 1980.

[159] A. M. et al. Sirunyan. "A multi-dimensional search for new heavy resonances decaying to boosted WW, WZ or ZZ boson pairs in the dijet final state at $\sqrt{s}$ = 13 TeV". In: *The European Physical Journal C* 80.3 (2020). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-020-7773-5. URL: http://dx.doi.org/10.1140/epjc/s10052-020-7773-5.

[160] A. et al. Hayrapetyan. "The CMS Statistical Analysis and Combination Tool: Combine". In: *Computing and Software for Big Science* 8.1 (2023). ISSN: 2510-2044. DOI: 10.1007/s41781-024-00121-4. URL: http://dx.doi.org/10.1007/s41781-024-00121-4.

[161] CMS. "Lund Plane Reweighting for Jet Substructure Correction". In: *CERN Document Server* (2023). URL: https://cds.cern.ch/record/2866330.

[162] Aram Hayrapetyan et al. “A method for correcting the substructure of multiprong jets using the Lund jet plane”. In: (July 2025). arXiv: 2507.07775 [hep-ex].

[163] Frédéric A. Dreyer, Gavin P. Salam, and Grégory Soyez. “The Lund jet plane”. In: *Journal of High Energy Physics* 2018.12 (2018). ISSN: 1029-8479. DOI: 10.1007/jhep12(2018)064. URL: http://dx.doi.org/10.1007/JHEP12(2018)064.

[164] CMS Collaboration (Internal nly). *Jet W-tagging*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetWtagging. Accessed: 2025-07-28. 2025.

[165] CMS Collaboration. “Performance of the CMS Level-1 trigger in proton-proton collisions at s = 13TeV”. In: *J. Instrum.* 15 (2020), P10017. DOI: 10.1088/1748-0221/15/10/P10017.

[166] CMS Collaboration (Internal only). *L1ECAL Prefiring Weight Recipe*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1ECALPrefiringWeightRecipe. Accessed: 2025-07-28. 2025.

[167] A. M. et al. Sirunyan. “Precision luminosity measurement in proton–proton collisions at

$$\sqrt{s} = 13\,\text{TeV}$$

in 2015 and 2016 at CMS”. In: *The European Physical Journal C* 81.9 (). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09538-2. URL: http://dx.doi.org/10.1140/epjc/s10052-021-09538-2.

[168] Joscha Knolle. *Precision luminosity measurement with proton-proton collisions at the CMS experiment in Run 2*. Tech. rep. Geneva: CERN, 2021. DOI: 10.22323/1.390.0748. URL: https://cds.cern.ch/record/2792593.

[169] Angela Giraldi. *Precision luminosity measurement with proton-proton collisions at the CMS experiment in Run 2*. 2022. arXiv: 2208.08214 [hep-ex]. URL: https://arxiv.org/abs/2208.08214.

[170] V. Khachatryan et al. “Measurement of the top quark mass using charged particles in pp collisions at $\sqrt{s} = 8$ TeV”. In: *Phys. Rev. D* 93.9 (2016), p. 092006. DOI: 10.1103/PhysRevD.93.092006. URL: https://doi.org/10.1103/PhysRevD.93.092006.

[171] F. James and M. Roos. “Minuit - a system for function minimization and analysis of the parameter errors and correlations”. In: *Computer Physics Communications* 10 (1975), pp. 343–367. DOI: 10.1016/0010-4655(75)90039-9.

[172] Eilam Gross and Ofer Vitells. “Trial factors for the look elsewhere effect in high energy physics”. In: *The European Physical Journal C* 70 (2010), pp. 525–530. DOI: 10.1140/epjc/s10052-010-1470-8. arXiv: arXiv:1005.1891 [physics.data-an].

[173] P. A. M. Dirac. *Principles of Quantum Mechanics*. Oxford University Press, 1939.

[174] John von Neumann. *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, 1955.

[175] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. 10th Anniversary Edition. Cambridge University Press, 2010.

[176] Peter W. Shor. “Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer”. In: *SIAM Journal on Computing* 26.5 (1997), pp. 1484–1509.

[177] Lov K. Grover. "A fast quantum mechanical algorithm for database search". In: *Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC '96)*. 1996, pp. 212–219.

[178] Richard P. Feynman. "Simulating physics with computers". In: *International Journal of Theoretical Physics* 21 (1982), pp. 467–488.

[179] Seth Lloyd. "Universal quantum simulators". In: *Science* 273.5278 (1996), pp. 1073–1078. DOI: 10.1126/science.273.5278.1073.

[180] Maria Schuld and Francesco Petruccione. "Machine learning with quantum computers". In: *Springer Series in the Sciences* (2021).

[181] Jonathan Romero, Jonathan P Olson, and Alan Aspuru-Guzik. "Quantum autoencoders for efficient compression of quantum data". In: *Quantum Science and Technology* 2.4 (Aug. 2017), p. 045001. DOI: 10.1088/2058-9565/aa8072. URL: https://dx.doi.org/10.1088/2058-9565/aa8072.

[182] Adriano Barenco, Charles H. Bennett, and Richard *et al.* Cleve. "Elementary gates for quantum computation". In: *Physical Review A* 52 (1995), pp. 3457–3467. DOI: 10.1103/PhysRevA.52.3457.

[183] James Stokes et al. "Quantum Natural Gradient". In: *Quantum* 4 (May 2020), p. 269. ISSN: 2521-327X. DOI: 10.22331/q-2020-05-25-269. URL: https://doi.org/10.22331/q-2020-05-25-269.

[184] A. Blance et al. "Quantum computing for high energy physics: State of the art and challenges". In: *SciPost Phys. Proc.* 7 (2022), p. 024. DOI: 10.21468/SciPostPhysProc.7.024. arXiv: 2112.01984 [quant-ph].

[185] Wen Guan et al. "Quantum machine learning in high energy physics". In: *Machine Learning: Science and Technology* 2.1 (Mar. 2021), p. 011003. ISSN: 2632-2153. DOI: 10.1088/2632-2153/abc17d. URL: http://dx.doi.org/10.1088/2632-2153/abc17d.

[186] M. Cerezo et al. "Variational quantum algorithms". In: *Nature Reviews Physics* 3 (2021), pp. 625–644.

[187] Alberto Di Meglio et al. "Quantum Computing for High-Energy Physics: State of the Art and Challenges". In: *PRX Quantum* 5 (3 Aug. 2024), p. 037001. DOI: 10.1103/PRXQuantum.5.037001. URL: https://link.aps.org/doi/10.1103/PRXQuantum.5.037001.

[188] Yi-An Chen and Kai-Feng Chen. "Jet discrimination with a quantum complete graph neural network". In: *Physical Review D* 111.1 (Jan. 2025). ISSN: 2470-0029. DOI: 10.1103/physrevd.111.016020. URL: http://dx.doi.org/10.1103/PhysRevD.111.016020.

[189] Navin Khaneja, Roger Brockett, and Steffen J. Glaser. "Time optimal control in spin systems". In: *Physical Review A* 63 (2001), p. 032308. DOI: 10.1103/PhysRevA.63.032308.

[190] J. J. Sakurai. *Modern Quantum Mechanics*. Revised. Addison–Wesley, 1994.

[191] PennyLane Team. *qml.AngleEmbedding — PennyLane 0.41.1 documentation*. Accessed: 2025-05-13. 2025. URL: https://docs.pennylane.ai/en/stable/code/api/pennylane.AngleEmbedding.html.

[192] PennyLane Team. *qml.AmplitudeEmbedding — PennyLane 0.41.1 documentation*. Accessed: 2025-05-13. 2025. URL: https://docs.pennylane.ai/en/stable/code/api/pennylane.AmplitudeEmbedding.html.

[193] Harry Buhrman et al. “Quantum Fingerprinting”. In: *Physical Review Letters* 87.16 (Sept. 2001). ISSN: 1079-7114. DOI: 10.1103/physrevlett.87.167902. URL: http://dx.doi.org/10.1103/PhysRevLett.87.167902.

[194] Jacob Biamonte et al. “Quantum machine learning”. In: *Nature* 549.7671 (2017), pp. 195–202. DOI: 10.1038/nature23474.

[195] Jack Y. Araz and Michael Spannowsky. *The role of data embedding in quantum autoencoders for improved anomaly detection*. 2024. arXiv: 2409.04519 [quant-ph]. URL: https://arxiv.org/abs/2409.04519.

[196] John Preskill. “Quantum Computing in the NISQ era and beyond”. In: *Quantum* 2 (2018), p. 79. DOI: 10.22331/q-2018-08-06-79.

[197] Jarrod R. McClean et al. “Barren plateaus in quantum neural network training landscapes”. In: *Nature Communications* 9.1 (2018), p. 4812. DOI: 10.1038/s41467-018-07090-4. arXiv: 1803.11173 [quant-ph]. URL: https://doi.org/10.1038/s41467-018-07090-4.

[198] Alessio Gianelle et al. “Quantum Machine Learning for b-jet charge identification”. In: *Journal of High Energy Physics* 2022.8 (Aug. 2022). ISSN: 1029-8479. DOI: 10.1007/jhep08(2022)014. URL: http://dx.doi.org/10.1007/JHEP08(2022)014.

[199] Vasilis Belis et al. “Quantum anomaly detection in the latent space of proton collision events at the LHC”. In: *Communications Physics* 7.1 (Oct. 2024). ISSN: 2399-3650. DOI: 10.1038/s42005-024-01811-6. URL: http://dx.doi.org/10.1038/s42005-024-01811-6.

[200] Aritra Bal et al. *1 Particle - 1 Qubit: Particle Physics Data Encoding for Quantum Machine Learning*. 2025. arXiv: 2502.17301 [hep-ph]. URL: https://arxiv.org/abs/2502.17301.

[201] Raghav Kansal et al. *JetNet*. Version 1.0. Zenodo, May 2021. DOI: 10.5281/zenodo.5502543. URL: https://doi.org/10.5281/zenodo.5502543.

[202] Gregor Kasieczka et al. “The Machine Learning landscape of top taggers”. In: *SciPost Physics* 7.1 (July 2019). ISSN: 2542-4653. DOI: 10.21468/scipostphys.7.1.014. URL: http://dx.doi.org/10.21468/SciPostPhys.7.1.014.

[203] Patrick Komiske, Eric Metodiev, and Jesse Thaler. *Pythia8 Quark and Gluon Jets for Energy Flow*. Version v1. Zenodo, May 2019. DOI: 10.5281/zenodo.3164691. URL: https://doi.org/10.5281/zenodo.3164691.

[204] S. Ovyn, X. Rouby, and V. Lemaitre. *Delphes, a framework for fast simulation of a generic collider experiment*. 2010. arXiv: 0903.2225 [hep-ph]. URL: https://arxiv.org/abs/0903.2225.

[205] Oz Amram et al. *Aspen Open Jets: Unlocking LHC Data for Foundation Models in Particle Physics*. 2024. arXiv: 2412.10504 [hep-ph]. URL: https://arxiv.org/abs/2412.10504.

[206] CMS Collaboration. *JetHT primary dataset in MINIAOD format from RunG of 2016 (/JetHT/Run2016G-UL2016_MiniAODv2-v2/MINIAOD)*. CERN Open Data Portal. Cite as: CMS Collaboration (2024). 2024. DOI: 10.7483/OPENDATA.CMS.1KTG.X0W4. URL: https://opendata.cern.ch/record/30508.

[207] CMS Collaboration. *JetHT primary dataset in MINIAOD format from RunH of 2016 (/JetHT/Run2016H-UL2016_MiniAODv2-v2/MINIAOD)*. CERN Open Data Portal. Cite as: CMS Collaboration (2024). 2024. DOI: 10.7483/OPENDATA.CMS.LT9E.T7RQ. URL: https://opendata.cern.ch/record/30541.

[208] Ville Bergholm et al. *PennyLane: Automatic differentiation of hybrid quantum-classical computations*. 2018. arXiv: 1811.04968 [quant-ph]. URL: https://arxiv.org/abs/1811.04968.

[209] Dougal Maclaurin, David Duvenaud, and Ryan P Adams. "Autograd: Effortless gradients in numpy". In: *ICML 2015 AutoML Workshop*. Vol. 238. 2015, p. 5.

[210] Melik Oughton. "1 Particle - 1 Qubit: Particle Physics Data Encoding for Quantum Machine Learning". Master's thesis. Imperial College London, Apr. 2025.

[211] Eric Pezone. "Quantum Machine Learning for Identifying Anomalous Hadronic Jets". Master's thesis. Imperial College London, Apr. 2025.

[212] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. URL: http://www.deeplearningbook.org.

[213] Eric M. Metodiev, Benjamin Nachman, and Jesse Thaler. "Classification without labels: learning from mixed samples in high energy physics". In: *Journal of High Energy Physics* 2017.10 (Oct. 2017). ISSN: 1029-8479. DOI: 10.1007/jhep10(2017)174. URL: http://dx.doi.org/10.1007/JHEP10(2017)174.

[214] J. Neyman and E.S. Pearson. "On the Problem of the Most Efficient Tests of Statistical Hypotheses". In: *Philosophical Transactions of the Royal Society of London. Series A* 231 (1933), pp. 289–337. DOI: 10.1098/rsta.1933.0009. URL: http://www.jstor.org/stable/91247.

[215] Oz Amram and Cristina Mantilla Suarez. "Tag N' Train: a technique to train improved classifiers on unlabeled data". In: *Journal of High Energy Physics* 2021.1 (Jan. 2021). ISSN: 1029-8479. DOI: 10.1007/jhep01(2021)153. URL: http://dx.doi.org/10.1007/JHEP01(2021)153.

[216] Anna Hallin et al. "Classifying anomalies through outer density estimation". In: *Physical Review D* 106.5 (Sept. 2022). ISSN: 2470-0029. DOI: 10.1103/physrevd.106.055006. URL: http://dx.doi.org/10.1103/PhysRevD.106.055006.

[217] Sang Eon Park et al. "Quasi anomalous knowledge: searching for new physics with embedded knowledge". In: *Journal of High Energy Physics* 2021.6 (June 2021). ISSN: 1029-8479. DOI: 10.1007/jhep06(2021)030. URL: http://dx.doi.org/10.1007/JHEP06(2021)030.

A. Appendix: Computing

A.1. Backpropagation Algorithm

The backpropagation algorithm [78, 212] is the cornerstone of training NNs. It efficiently computes the gradient of a scalar loss function $\mathcal{L}$ with respect to every parameter (weights $W^{(\ell)}$ and biases $\mathbf{b}^{(\ell)}$) by repeatedly applying the chain rule. Those gradients are then used by an optimiser routine such as Adam to update the parameters by minimising $\mathcal{L}$.

Consider a feed-forward NN with L layers. We adopt the column-vector convention whereby activations $\mathbf{a}^{(\ell)} \in \mathbb{R}^{n_\ell}$ and weight matrices $W^{(\ell)} \in \mathbb{R}^{n_\ell \times n_{\ell-1}}$. Backpropagation proceeds in four steps:

- **Forward pass**: propagate the input $\mathbf{x}$ through the network and *cache* the pre-activations $\mathbf{z}^{(\ell)}$ and activations $\mathbf{a}^{(\ell)}$:

$$\mathbf{a}^{(0)} = \mathbf{x}, \tag{A.1}$$
$$\mathbf{z}^{(\ell)} = W^{(\ell)}\mathbf{a}^{(\ell-1)} + \mathbf{b}^{(\ell)}, \qquad \ell = 1, \dots, L, \tag{A.2}$$
$$\mathbf{a}^{(\ell)} = \phi^{(\ell)}\big(\mathbf{z}^{(\ell)}\big), \qquad \ell = 1, \dots, L, \tag{A.3}$$

  where $\phi^{(\ell)}$ is the element-wise activation function of layer ℓ.

- **Backward pass**: propagate the derivative term $\boldsymbol{\delta}^{(\ell)} \equiv \partial\mathcal{L}/\partial\mathbf{z}^{(\ell)}$ from the output layer back to the input:

$$\boldsymbol{\delta}^{(L)} = \nabla_{\mathbf{a}^{(L)}}\mathcal{L} \;\odot\; \phi'^{(L)}\big(\mathbf{z}^{(L)}\big), \tag{A.4}$$
$$\boldsymbol{\delta}^{(\ell)} = \big(W^{(\ell+1)}\big)^{\top}\boldsymbol{\delta}^{(\ell+1)} \;\odot\; \phi'^{(\ell)}\big(\mathbf{z}^{(\ell)}\big), \qquad \ell = L-1, \dots, 1. \tag{A.5}$$

- **Gradient computation**: for each layer, compute as follows -

$$\frac{\partial\mathcal{L}}{\partial W^{(\ell)}} = \boldsymbol{\delta}^{(\ell)}\,\big(\mathbf{a}^{(\ell-1)}\big)^{\top}, \tag{A.6}$$
$$\frac{\partial\mathcal{L}}{\partial \mathbf{b}^{(\ell)}} = \boldsymbol{\delta}^{(\ell)}, \tag{A.7}$$

  which, for a mini-batch of size m, is then averaged over the m examples.

- **Parameter update**: e.g. stochastic-gradient descent with learning rate η:

$$W^{(\ell)} \leftarrow W^{(\ell)} - \eta\,\frac{\partial\mathcal{L}}{\partial W^{(\ell)}}, \qquad \mathbf{b}^{(\ell)} \leftarrow \mathbf{b}^{(\ell)} - \eta\,\frac{\partial\mathcal{L}}{\partial \mathbf{b}^{(\ell)}}. \tag{A.8}$$

These steps are then repeated for each epoch until convergence or a stopping criterion is met.

B. Appendix: CMS-EXO-22-026

B.1. Trigger Efficiency Plots

The triggers used to select events have different efficiencies depending on the year the data was recorded. The goal is to select interesting events containing highly energetic jets as can be seen in the trigger efficiency distributions of Figure B.1 as a function of m_{jj} and jet soft-drop mass m_{SD}

B.2. Signal Injection Bias Studies

To validate the robustness of the VAE-QR methodology and quantify its sensitivity, comprehensive bias tests are performed using the mock MC QCD dataset with an integrated luminosity of 26.8 fb^{-1}. The tests involve injecting signals of increasing cross-section into the dataset used for the QR training (on MC) to verify exactly two crucial assumptions:

- the background shape remains unaffected by the presence of signal
- subsequent fits can recover the signal that was injected (see Figures B.3a,B.3b and B.3c)

The benchmark signal model employed is an $X \rightarrow YY'$ resonance with $M_X = 2$ TeV, and two configurations of daughter particle masses: a more boosted scenario with $M_Y = 80$ GeV, $M_{Y'} = 170$ GeV, and a less boosted scenario with $M_Y = 400$ GeV, $M_{Y'} = 170$ GeV. The significance of the signal is evaluated across a range of injected cross-sections, with p-value scans shown separately for each anomaly category as well as for their combination.

The results shown in Figure B.2 demonstrate that the k-fold procedure described in 5.6.2.1 is able to mitigate any potential biases that could arise from the presence of signal. It also makes a convincing case for the necessity of the three-category limit setting procedure compared to both the inclusive dijet fit and a single-category analysis. It is seen that for the less boosted signal configuration ($M_Y = 400$ GeV, $M_{Y'} = 170$ GeV), the jets from the decay are less likely to merge, creating more complex event topologies that are captured differently across the anomaly categories. The tighter quantile regions preferentially select events where the two jets exhibit greater merging, resulting in a sharper signal shape compared to the looser selections. This difference in responses across categories justifies the limit setting procedure described in Section 5.8.8, as it enables the analysis to substantially enhance the overall statistical power of the search. The corresponding fits for the three categories used in the limit setting procedure are shown in Figures B.3a,B.3b and B.3c.

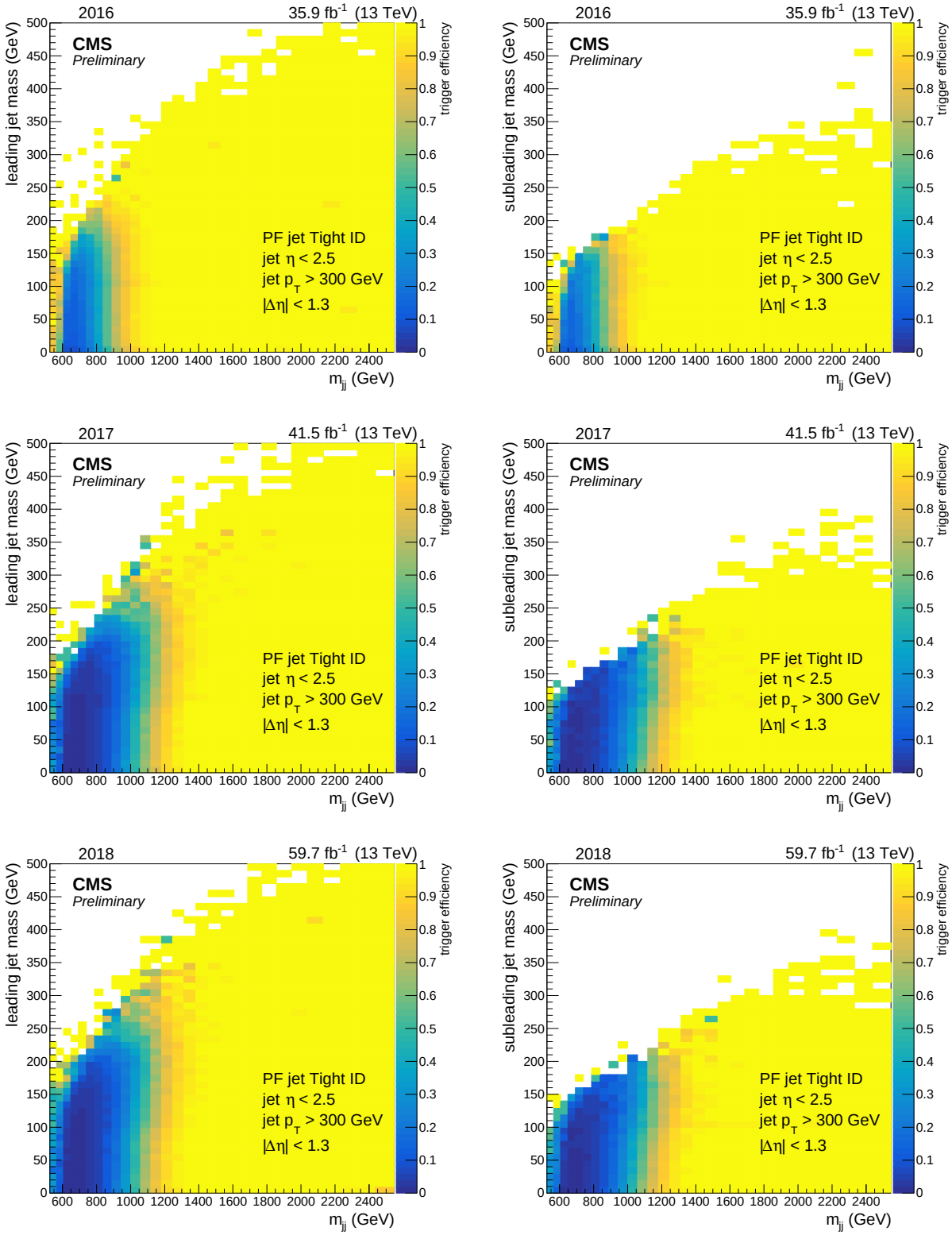


Figure B.1.: Trigger Efficiency for each year as a function of m_{jj} and either leading (left) or subleading (right) jet mass

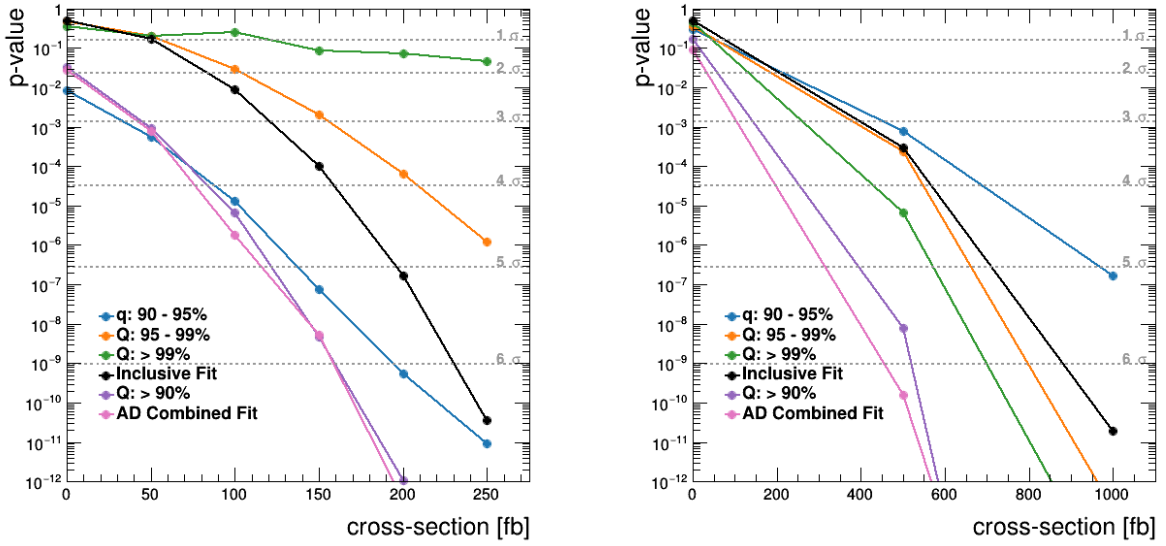
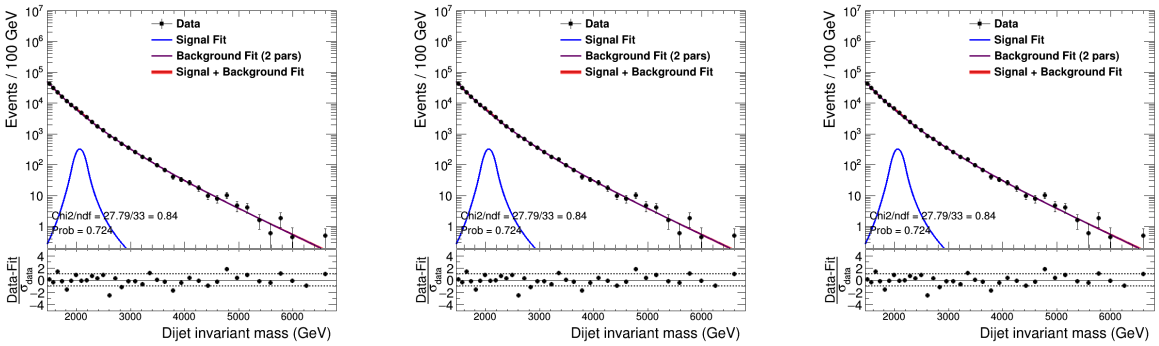


Figure B.2.: p-value scans for the injected signal $M_X = 2$ TeV with two different decay product masses $M_Y = 80$ GeV, $M_{Y'} = 170$ GeV (left) and $M_Y = 400$ GeV, $M_{Y'} = 170$ GeV (right)



(a) Signal+Background Fit for the events lying in the category between the 90$^{\text{th}}$ and 95$^{\text{th}}$ percentiles. The aforementioned signal model being injected into the MC QCD dataset with a cross-section of 0.25 pb corresponding to a expected significance of about 6.5σ

(b) Signal+Background Fit for the events lying in the category between the 95$^{\text{th}}$ and 99$^{\text{th}}$ percentiles.

(c) Signal+Background Fit for the events lying in the category above the 99$^{\text{th}}$ percentile

Figure B.3.: Signal + Background fits in the three categories, with an injected signal

B.3. Other methods in CMS-EXO-22-026

The analysis CMS-EXO-22-026 contained a total of 5 anomaly detection methods, out of which the VAE-QR method is the focus of this thesis. The other methods are described briefly in this Appendix.

The first three methods: CWoLa Hunting, TNT and CATHODE follow the paradigm of weak supervision, which must be understood first. In the supervised paradigm, one possesses clearly labelled datasets where the nature of each event, whether signal or background, is exactly known. On the other hand, one could instead have two datasets, as shown in Figure B.4, where the exact nature of individual events is unknown, but the relative proportions of signal and background differ between the two datasets.

For weak supervision in the classification context, it is not necessary to know the exact proportions of background and signal in the datasets. It is only required that there are two mixed datasets with different fractions of signal and background [213]. In the fully supervised case, this reduces to a special case where one dataset contains 100% signal events and another contains 100% background events, and the network learns to distinguish between these pure samples.

According to the Neyman-Pearson lemma, the optimal classifier for distinguishing between two hypotheses is the likelihood ratio [214]. It can be demonstrated that the optimal classifier for distinguishing between two mixed datasets with different signal-background compositions is monotonically related to the optimal classifier for distinguishing between pure signal and background samples [213]. This means that in the limit of sufficient statistics in the mixed samples, a classifier trained to distinguish between mixed datasets can converge to the performance of a fully supervised classifier.

In the particle physics context, this approach is particularly valuable as the available data can be split into a signal-enriched region (SR), and a signal-depleted sideband. A classifier can be trained to distinguish between these two regions, and under appropriate conditions, this would optimally converge to the same performance as a classifier trained to distinguish between unambiguously labelled signal and background events. This approach has proven effective in data-driven analyses, where the usage of simulated signal samples is undesirable, as is often the case when searching for unknown physics phenomena.

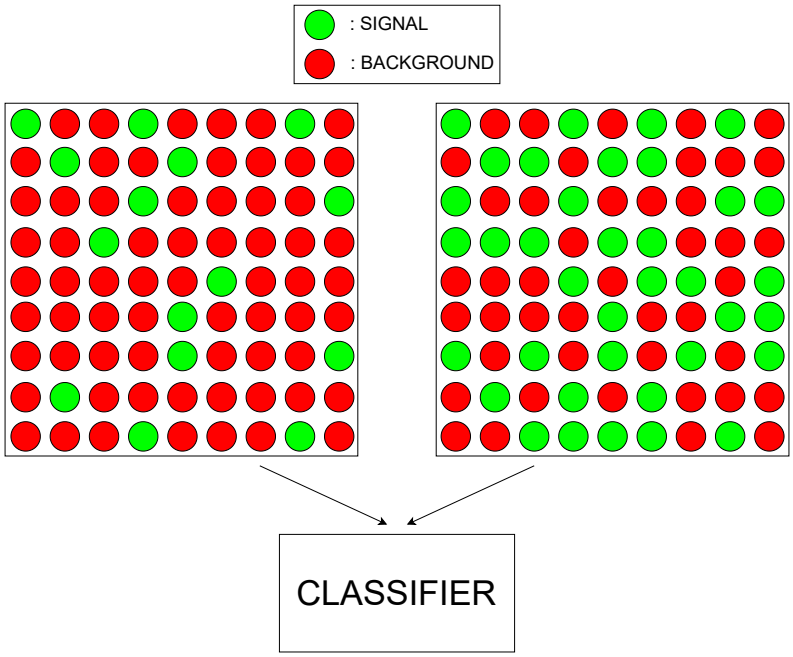


Figure B.4.: Illustration of the weak supervision paradigm. Instead of having purely labelled signal and background samples as in the supervised case, weak supervision utilises two mixed samples with different signal fractions. A classifier trained to distinguish between these mixed samples approximates a classifier trained on pure samples.

The remaining methods that are part of the analysis are now described below:

- **CWoLa Hunting**: This approach combines weakly supervised classifier training with the assumption of a resonant signal. A signal mass window in m_{jj} is selected to define a signal region (SR), while two sidebands on either side of that mass region serve as the

background region. A classifier is trained to distinguish between these two groups of events using the weakly supervised method. This procedure is repeated for different signal mass windows to cover the full m_{jj} range, as detailed in Table B.1. If a narrow-width resonant signal is present, it will be localised in a particular mass region, and the SR will contain the majority of the anomalous signal events compared to the sidebands. The classifier would then learn to tag these events.

An important consideration is that the features used for classification must not be correlated with m_{jj} to prevent sculpting of the mass distribution. This is done by reweighting events in the two sidebands of higher and lower mass respectively, to have equal weights. On top of that, events in the signal region are reweighted to be of equal weight as the sidebands. The input features used in the CWoLa Hunting approach include:

- m_{SD}: Soft-drop mass of the jet
- $\tau_{21} \rightarrow$ N-subjettiness ratio τ_2/τ_1
- $\tau_{32} \rightarrow$ N-subjettiness ratio τ_3/τ_2
- $\tau_{43} \rightarrow$ N-subjettiness ratio τ_4/τ_3
- n_{PF}: Number of PFCands in the jet
- $LSF_3 \rightarrow$: Les Houches angularity variable
- DeepB: B-tagging discriminator

The CWoLa Hunting approach used in this analysis employs two per-jet classifiers instead of an event classifier, with these classifier scores being combined thereafter to obtain a final event-level anomaly score.

- **TNT**: This method specifically targets signals where both jets in the event are anomalous, and operates through an autoencoder-based preselection strategy. The same features and signal region plus sideband definitions (as detailed in Table B.1) are utilised for the TNT method as were employed for CWoLa Hunting.

  In TNT, as for CWoLa hunting, two jet-level autoencoders are trained. For each event, the two jets are labelled as J_1 and J_2. Unlike the original implementation [215] where the heavier and lighter jets were explicitly identified, in this analysis the J_1 and J_2 labels are randomly assigned to the two jets in each event. The autoencoders are trained on the m_{jj} sidebands, for each considered signal region, and then used to compute reconstruction scores for each jet, with higher scores indicating potential anomalies.

  The jets can then be labeled as background-like or signal-like, based on what percentile of the autoencoder reconstruction score they lie in. Specifically, jets from each signal region with reconstruction scores in the top 20% are classified as signal-like, while those in the m_{jj} sideband, or in the SR with scores in the bottom 40% are classified as background-like. These jets are then combined to create signal-enriched and signal-depleted jet datasets, which are subsequently used to train a weak classifier, following a similar approach to that used in CWoLa Hunting.

- **CATHODE**: The primary goal is to learn the likelihood ratio between signal and background probability distributions, defined as:

$$R(\vec{x}) = \frac{p_{\text{sig+bkg}}(\vec{x})}{p_{\text{bkg}}(\vec{x})} \tag{B.1}$$

  CATHODE [216] employs a sophisticated density estimation technique to construct a background model for each signal region, using the background density in the sidebands. In the first step, it uses normalising flows to learn the background density distribution. This is achieved by mapping the background density in the sideband (SB) to a normalising flow probability density, which can smoothly transform between different probability distributions.

  The following input features are used:

  - Leading jet mass m_{j1}
  - Sub-leading jet mass m_{j2}
  - Mass difference $\Delta m_{j1j2} = |m_{j1} - m_{j2}|$
  - N-subjettiness ratio τ_{41} for both jets

  Having learnt $p_{\text{bkg}}(x|m_{\text{jj}} \notin \text{SR})$, events can be sampled to obtain $p_{\text{bkg}}(x|m_{\text{jj}} \in \text{SR})$. Thereafter, a classifier can be trained to distinguish between these synthetic events and actual events in the signal region which have the underlying distribution $p_{\text{sig+bkg}}(x|m_{\text{jj}} \in \text{SR})$, effectively creating an anomaly detection metric, as was done for CWoLa Hunting and TNT. An additional version of CATHODE, named CATHODE-b that uses the b-tagging DeepCSV score of each leading jet is also included in the analysis, and is observed to have higher sensitivity to processes which produce b-jets.

- **QUAK**: This can best be described as a semi-supervised method, since it requires the usage of signal and background MC simulation for training. The QUAK anomaly detection method [217] utilises supervised training directly on simulated signals and background samples without intermediate generative models. A total of 6 normalizing flows are trained on a mixture of Monte Carlo signal simulations, with the following masses of the decay products B and C for the decay $A \rightarrow BC$: (80,80), (80,170), (80,400), (170,170), (170,400), and (400,400) GeV. Input features include: $\rho = \frac{m_1}{p_{\text{T}}}$, n-subjettiness ratios τ_{21}, τ_{32}, and τ_{43}, a modified subjettiness ratio $\tau_s = \frac{\sqrt{m_{21}}}{m_1}$, the DeepB tagger score, and the number of jet constituent PFCands (n_{PF}). Another normalising flow is then trained to learn the background MC distribution. QUAK produces a two-dimensional anomaly score using the negative log-likelihood losses of the normalising flow trained on background and the combination of those trained on the mixture of signal priors (using an L5 norm). Prior to performing a selection, the losses are decorrelated from m_{jj} to ensure they have an approximately Gaussian distribution. Thereafter, for each mass hypothesis m_{H} that is tested, an SR and a sideband is defined, with the sideband serving as the background template. Finally, events are selected using a background-excluding contour constructed by binning this 2D loss space using polar coordinates, and selecting the least populated bins in the SR iteratively until a specified number of events have been selected. These selected events can then be used for the common bump-hunt and limit setting procedure.

Table B.1.: Signal regions used by the weakly supervised methods, the resonance masses considered with each bin, the efficiency of the anomaly score cut, and the number of data events in the bin after pre-selection (before any anomaly score cut).

Bin Name	Range (GeV)	Eff. Cut	Signal Masses (GeV)	Num. data events
A0	1350-1650	-	-	13.8M
A1	1650-2017	1%	1800, 1900	4.5M
A2	2017-2465	1%	2200, 2300	1.4M
A3	2465-3013	1%	2600, 2700, 2800	400k
A4	3013-3682	3%	3200, 3300, 3400, 3500	100k
A5	3682-4500	3%	3900, 4100, 4200, 4300	22k
A6	4500-5500	5%	4800, 4900, 5000, 5100, 5200	3.9k
A7	5500-8000	-	-	479
B0	1492-1824	-	-	6.6M
B1	1824-2230	1%	2000, 2100	2.1M
B2	2230-2725	1%	2400, 2500	630k
B3	2725-3331	1%	2900, 3000, 3100	170k
B4	3331-4071	3%	3600, 3700, 3800	42k
B5	4071-4975	3%	4400, 4500, 4600, 4700	8.5k
B6	4975-6081	5%	5300, 5400, 5500, 5600, 5700, 5800	1.3k
B7	6081-8000	-	-	144

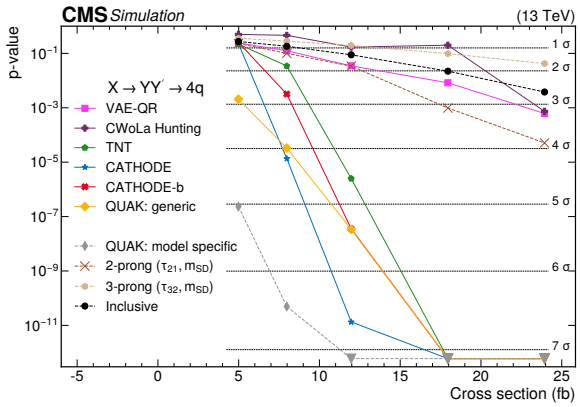
In the case of QUAK, it is important to note that the mixture of signal priors that it is trained on, may also contain the signal on which this method is later used to set limits on. While the limits obtained using QUAK are competitive with the other methods and better than the inclusive search with no selection, a degradation of approximately 30% is observed when a given signal model is removed from the training prior of QUAK, before subsequently using it to set limits on that model.

B.3.1. Expected Sensitivity Studies

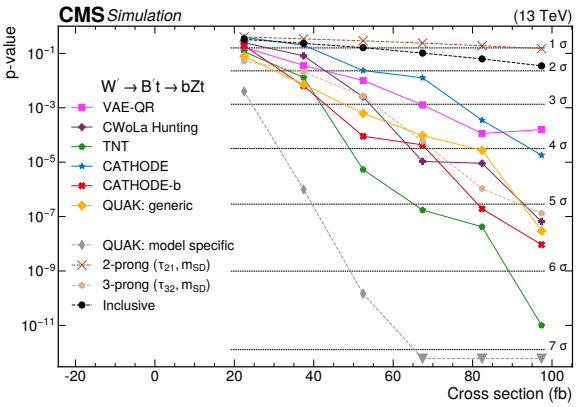
To assess the potential of the proposed analysis methodologies, comprehensive expected sensitivity studies are performed. Whilst all methods in the analysis are expected to be sensitive to a wide range of topologies, a dedicated study is carried out using the mock MC dataset corresponding to an integrated luminosity of 26.8fb^{-1}.

Two such studies are conducted, each utilising different hypothetical signal models injected into the background MC dataset with gradually increasing cross-sections. The first signal model is the $X \rightarrow YY' \rightarrow 4q$, characterised by a 2 + 2 prong structure, and the second is the $W' \rightarrow B't \rightarrow bZt$ decay, featuring a 3 + 3 prong topology. These complementary topologies were selected to demonstrate the broad applicability of the analysis methods across different final state configurations commonly predicted by theories beyond the SM.

For each signal model, the complete search procedure was implemented, including the characteristic search for resonant anomalies over the m_{jj} spectrum. The expected statistical significance for each signal model was calculated using the methodology highlighted in Section 5.8. The results of these studies are presented in Figure B.5, which illustrates the expected discovery potential for both signal topologies across the range of cross-sections investigated.



(a) Expected statistical significance for the $X \to YY' \to 4q$ signal model as a function of injected cross-section. The horizontal dashed lines indicate the various $N\,\sigma$ thresholds, with discovery typically being claimed at 5σ.



(b) Expected statistical significance for the $W' \to B't \to bZt$ signal model as a function of injected cross-section. The sensitivity curve demonstrates the enhanced discovery potential for this 3 + 3 prong topology.

Figure B.5.: Expected sensitivity studies for two representative signal models. Both panels show the statistical significance as a function of the injected signal cross-section, demonstrating the discovery potential of the various methods across different topologies.

This study demonstrates that while the inclusive search procedure without any selection exhibits some sensitivity to the presence of signal, all methods employed in this analysis demonstrate significantly enhanced sensitivity compared to the inclusive approach. This improvement in discovery potential validates and justifies the usage of these sophisticated ML techniques for new physics searches and demonstrates their ability to extract subtle signal features from complex background environments.

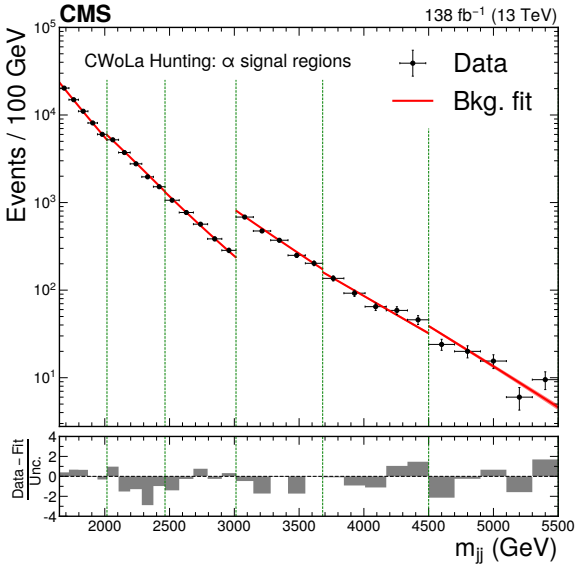
The optimal performance was achieved, as expected, through the implementation of a specially designed model-specific version of QUAK trained on a signal prior consisting exclusively of the signal model upon which limits are subsequently set. This configuration corresponds to a fully supervised search procedure and serves as a benchmark against which the model-agnostic methods that make up this analysis can be evaluated. These sensitivity studies thus provide confidence that the proposed methods in CMS-EXO-22-026 offer substantial discovery potential across a spectrum of potential new physics scenarios while preserving the model-agnosticism required for broad searches at the LHC.

B.3.2. Background-only Fits

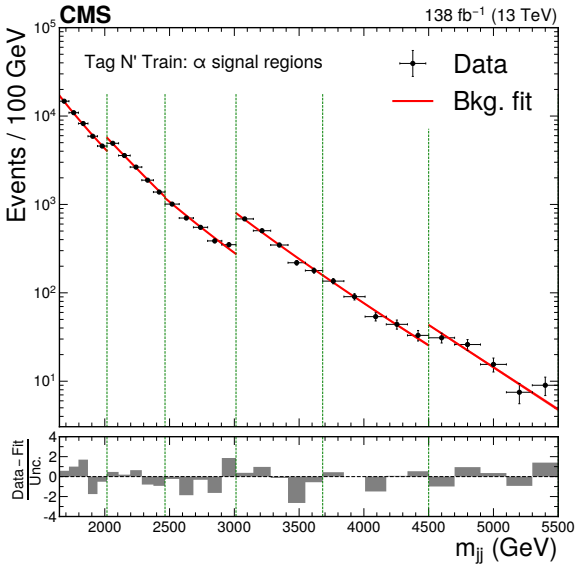
This Appendix presents the background-only fits for all methods other than the VAE-QR, whose background-only fit was previously shown in Figure 5.26.

The VAE-QR approach has the advantage in its ability to perform the background fit across the entire m_{jj} spectrum simultaneously. In contrast, the remaining methods in this analysis, namely CWoLa Hunting, TNT, CATHODE, and QUAK, define and use narrow sliding m_{jj} windows for signal searches, as detailed in this Appendix and listed in Table B.1. As a result, individual background-only fits are performed within each narrow mass interval, with each fit independently characterising the background behaviour within its respective window. To present a comprehensive view of the background modelling performance across the entire m_{jj} range, the results from these individual fits are plotted by stitching together the fits from each window, and shown in Figure B.6. These fits serve as the foundation upon which searches are

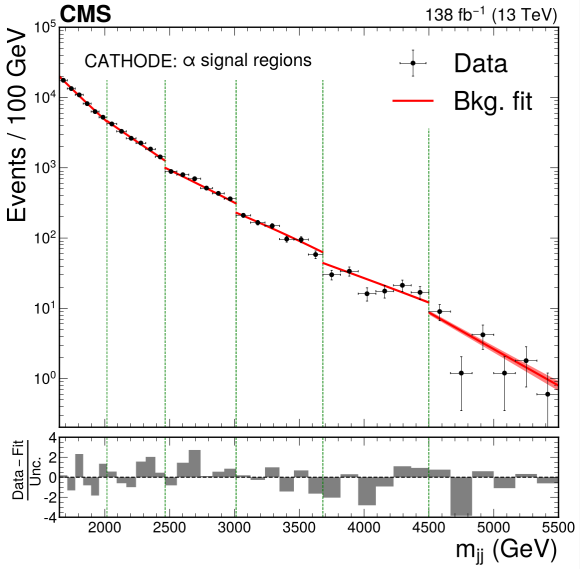
conducted, with any significant deviation potentially indicating the presence of new physics signatures.



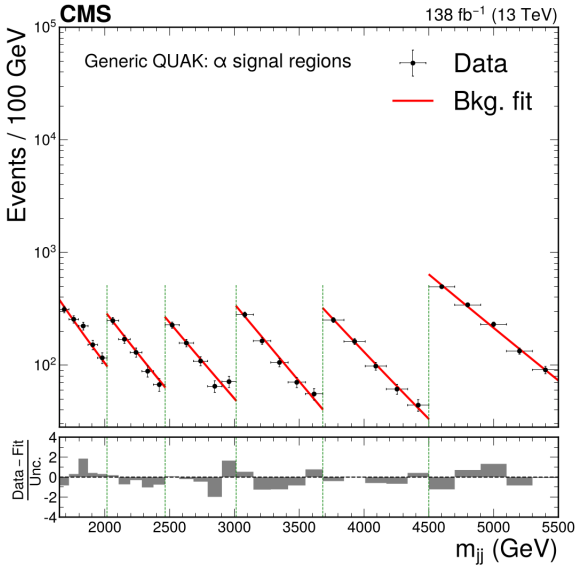
(a) CWoLa Hunting background-only fit results across the full m_{jj} spectrum.



(b) TNT background-only fits



(c) CATHODE background-only fits



(d) QUAK background-only fits

Figure B.6.: Background-only fit results for the other four anomaly detection methods. Each panel shows the background as obtained by stitching together individual fits from narrow m_{jj} windows

B.3.3. Comparison and Complementarity between Methods

The methods presented in this analysis share a common dataset and phase space region, yet they differ significantly in both their algorithmic design and feature selection. This methodological diversity naturally leads to variations in sensitivity across different signal topologies. This section examines these differences to establish the complementarity of the employed techniques.

Empirical observations indicate distinctive performance patterns across methods. The VAE-QR demonstrates particular sensitivity to signals where the daughter decay particles possess high masses. Conversely, other methods exhibit enhanced performance for signatures with specific jet substructure patterns or distinctive angular distributions.

To quantitatively assess the relative sensitivities of these methods, various benchmark signals were injected into the mock Monte Carlo dataset at different cross sections. For each method, the minimum cross section required to achieve a 3σ and 5σ expected significance was calculated. Figure B.7 illustrates these threshold cross sections across the different methods and benchmark signals, providing a comparative view of each method's sensitivity to specific signal topologies.

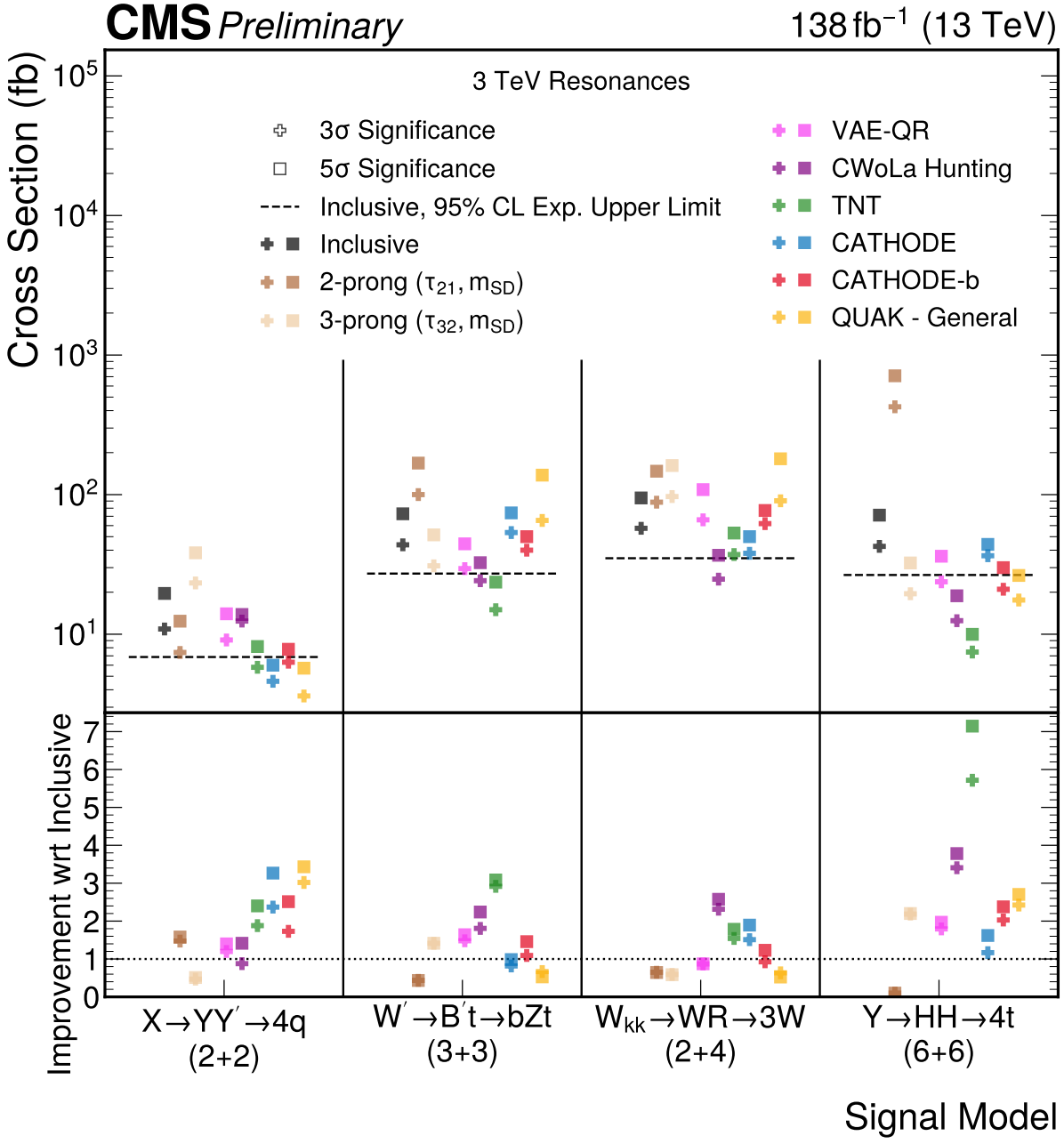
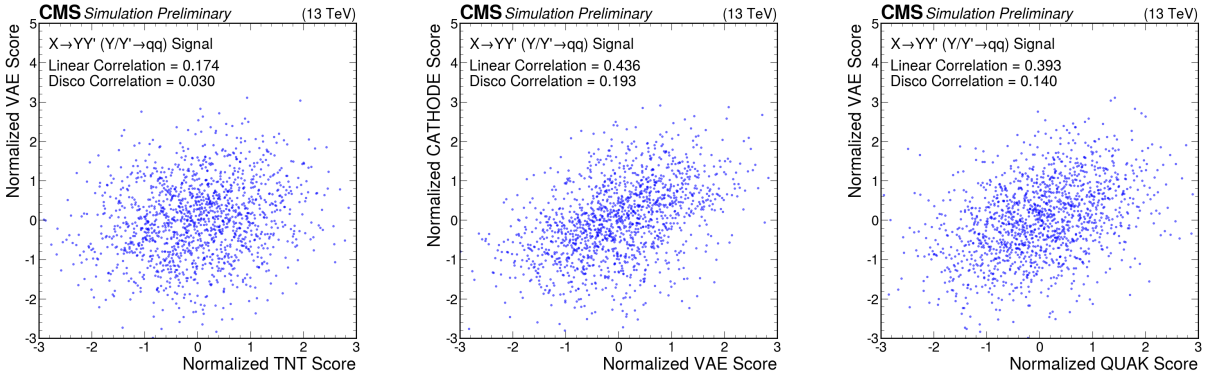


Figure B.7.: Injected cross sections of each benchmark signal model required to achieve a $3\sigma/5\sigma$ expected significance, using each anomaly detection method.

A more direct comparison was performed by computing normalised anomaly scores for each method on a single benchmark signal, specifically the $X \to YY' \to 4q$ process. These scores were normalised to the range $(0, 1)$ to facilitate direct comparison. Figure B.8 presents scatter plots comparing the VAE-QR scores against those from three other methods, while Figure B.9 displays the complete correlation matrix of Pearson correlation coefficients between each pair of method scores. Analysis of the correlation coefficients reveals moderate correlations between



(a) TNT vs VAE-QR **(b)** CATHODE vs VAE-QR **(c)** QUAK vs VAE-QR

Figure B.8.: Distribution of the normalised anomaly scores of 3 out of the 4 anomaly detection methods as compared to that of the VAE-QR, using the $X \to YY' \to 4q$ benchmark signal

certain method pairs, with the most pronounced correlation (> 0.5) observed between the TNT and CWoLa Hunting methods. This higher correlation is expected given that these methods share core principles and differ primarily in the application of an additional autoencoder-based preselection step in the TNT approach. Nevertheless, no method pair exhibits correlations sufficiently high to render any single method meaningless.

The overall low correlation between methods indicates that each captures different aspects of the signal topology. This complementarity is particularly valuable in model-agnostic searches where the exact nature of potential new physics signals remains unknown. Methods that prioritise different features or employ varied learning strategies can collectively provide broader coverage of the potential signal space than any single method alone.

The studies presented here demonstrate conclusively that the suite of methods presented in this analysis provide comprehensive coverage across a wide range of signal topologies. Each method contributes unique sensitivity to different aspects of potential signals, ensuring that the combined analysis maximises the probability of detecting anomalous physics processes, should they exist in the data. This complementarity is essential for robust model-agnostic searches in high-energy physics, where the precise characteristics of new physics remain mostly unknown.

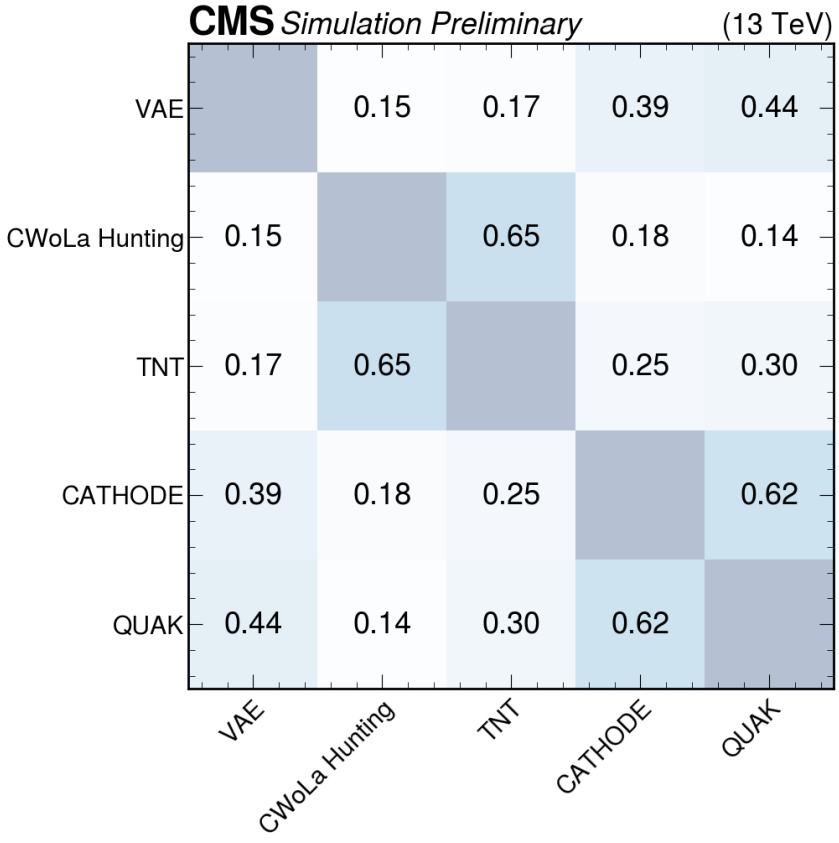


Figure B.9.: Correlation Matrix showing the Pearson coefficient of correlation between the anomaly scores obtained by each method

Acknowledgment

I would like to express my sincere gratitude to my supervisors, Prof. Dr. Markus Klute, and particularly Dr. Benedikt Maier, whose continuous guidance and invaluable support have been essential throughout my PhD journey. Without Dr. Maier's mentorship the last 2.5 years, my exploration of machine learning, both classical and quantum, would not have progressed this far.

My heartfelt thanks also go out to my co-referee, P.D. Dr. Roger Wolf, for his kind and constructive reviews, insightful comments, and also for making my integration into the institute smooth when I first arrived in Germany three years ago.

I specially thank Prof. Michael Spannowsky for his pivotal insights and guidance on Quantum Machine Learning, which significantly influenced the latter half of my thesis. I also appreciate the remarkable contributions and dedicated work of master's students Melik Oughton [210] and Eric Pezone [211] on the QML project. I also thank my collaborators within CMS, in particular Dr. Jennifer Ngadiuba, Dr. Oz Amram, Dr. Louis Moureaux, among many others, for their guidance during the CMS-EXO-22-026 analysis.

I extend warm thanks to my friends in Karlsruhe: Alejandro, Xunwu, Olha, Marcel, Johannes, Christian, Ralf, Cedric and many others at the ETP, whose continuous support, insightful physics discussions, impromptu coffee and tea break chats, and the last-minute thesis reviews greatly contributed to making these years at KIT memorable. I appreciate in particular the insightful comments I received on my thesis from Dr. Nils Faltermann and Dr. Brendan Regnery.

I am grateful to my friends from undergrad, Asimpunya, Dileep, Anirban, Arunava, Amatra, Ansuman, and Swastik, for our weekly physics discussions and exchanges on various academic topics. I thank Dr. A.K. Dubey, my undergraduate supervisor, whose encouragement led me to pursue doctoral studies, and Dr. G. Mukhopadhyay, who significantly influenced my decision to major in Physics. I also thank my friends from outside of the physics universe, Aditya, Abhishek, Nikhil, Yash, Rohan, Saksham, Adiraj, Aniket, Nidhi, and many others, whose friendship provided valuable perspective during my academic journey.

Finally and most importantly, I thank my partner Pankhuri, my parents Utpal and Paromita, my sister Aheli, and my aunt Sangita, whose unwavering support, encouragement, and belief in my abilities have been the cornerstone of my academic and personal career achievements.