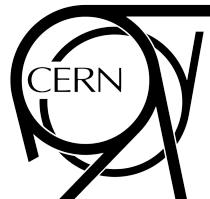


AN EXPLORATION OF DEEP LEARNING METHODS FOR ANOMALY DETECTION IN HIGH ENERGY JETS

JORDAN PFEIFER



BROWN



Bachelor of Sciences, Mathematical Physics
Bachelor of Arts, Computer Science
Department of Physics
Dr. Greg Landsberg and Dr. Loukas Gouskos
Brown University

April 2025

ABSTRACT

The search for Beyond the Standard Model (BSM) physics at the Large Hadron Collider (LHC) increasingly relies on model-agnostic anomaly detection techniques to identify deviations from Standard Model predictions. This thesis explores deep learning methods for unsupervised anomaly detection in jets using simulated proton-proton collision data at $\sqrt{s} = 13$ TeV. Jets are represented as sets of particle flow candidates and analyzed using several neural network architectures: fully-connected autoencoders, transformers, and graph autoencoders with EdgeConv layers. Models are trained on a QCD jet background and evaluated on their ability to identify W jets as signal-like anomalies.

Each model's performance is assessed via anomaly score, with Area Under the Receiver Operating Characteristic Curve (AUC) used as the primary metric. The supervised classification baseline achieves an AUC of 0.69, while unsupervised models attain AUCs in the range of 0.58–0.61. These results suggest limited separability between W and QCD jets in the chosen feature space and highlight the difficulty of detecting structured signals using reconstruction-based AD methods.

We identify several contributing factors to this performance: lack of discriminative input features, dataset similarity, and architectural bottlenecks. As such, extensive hyper-parameter tuning and feature engineering were deferred in favor of establishing baseline model behavior. The findings underscore the challenges of deploying unsupervised learning for jet anomaly detection and motivate future work in advanced representations, graph construction, and real-data application. This work contributes to a growing body of efforts to build robust, model-independent search pipelines for discovery at the LHC.

ACKNOWLEDGMENTS

Thank you to Profs. Greg Landsberg and Loukas Gouskos for their guidance and support over many months.

I'm especially grateful to Marko Stamenkovic for his advice, encouragement, and company during two summers at CERN.

Thanks to Rohan Pankaj and Ekin Secilmis for being thoughtful collaborators, and to Jason Huang, Gavin Pitt, and Hayden Miller for the office culture.

Finally, love and thanks to my friends and family.

CONTENTS

| | | |
|-------|------------------------------------------------------|----|
| 1 | INTRODUCTION | 1 |
| 1.1 | Beyond the Standard Model | 1 |
| 1.1.1 | The Hierarchy Problem | 1 |
| 1.1.2 | Dark Matter | 1 |
| 1.1.3 | Flavor Puzzles and Further Open Questions | 2 |
| 1.1.4 | Searching for BSM Physics | 2 |
| 1.2 | Challenges in Jet-Based Searches | 2 |
| 1.3 | Related Work | 3 |
| 1.3.1 | Benchmarking Datasets | 4 |
| 2 | JETS | 6 |
| 2.0.1 | Quantum Chromodynamics (QCD) Jets | 6 |
| 2.0.2 | W Boson Jets (W+jets) | 6 |
| 2.0.3 | Measuring Jets: The Detector | 7 |
| 3 | MACHINE LEARNING ARCHITECTURES FOR ANOMALY DETECTION | 9 |
| 3.0.1 | Autoencoders and Variants | 9 |
| 3.0.2 | Convolutional Autoencoders | 9 |
| 3.0.3 | Variational Autoencoders | 9 |
| 3.0.4 | Transformer-Based Models | 10 |
| 3.0.5 | Graph Autoencoders with EdgeConv | 10 |
| 4 | APPLICATION TO SIMULATED LHC DATA | 11 |
| 4.1 | Data Sample | 11 |
| 4.1.1 | Data Flow | 11 |
| 4.1.2 | Event Selection and Cutoff Criteria | 11 |
| 4.2 | Data Cleaning and Feature Engineering | 13 |
| 4.2.1 | Jet Level | 13 |
| 4.2.2 | Dataset Level | 14 |
| 4.3 | Data Structure | 14 |
| 4.3.1 | Array Format | 14 |
| 4.3.2 | Graph Format | 14 |
| 4.4 | Model Implementations | 18 |
| 4.4.1 | Autoencoder | 18 |
| 4.4.2 | Graph Autoencoder | 18 |
| 4.4.3 | Baseline: Supervised Learning/Classification | 19 |
| 4.5 | Training, Evaluation, and Tuning | 19 |
| 5 | RESULTS AND INTERPRETATION | 20 |
| 5.1 | Characterizing the Background Model | 20 |
| 5.2 | Interpretation and Comparison | 21 |
| 6 | DISCUSSION AND FUTURE WORK | 24 |
| 6.1 | Interpretation of Results | 24 |
| 6.2 | Future Work | 24 |
| | BIBLIOGRAPHY | 26 |

1.1 BEYOND THE STANDARD MODEL

The Standard Model (SM) of particle physics [1–3] accurately describes a wide range of high-energy physics processes and interactions [4]. In 2012, the discovery of the Higgs boson¹, confirmed the mechanism of electroweak symmetry breaking [6, 7]. Despite this success, the Standard Model is widely regarded as incomplete, with open questions including the hierarchy problem, the existence of dark matter, and various flavor puzzles.

1.1.1 *The Hierarchy Problem*

The SM does not explain the "unnatural" discrepancy between the mass of the Higgs boson (125 GeV) and the Planck scale (10^{19} GeV), which in turn affects the disparity in scale between the weak force and gravity [8, 9]. The quadratically divergent loop corrections to the Higgs mass leads us to believe the Higgs should measure around the Planck scale; however, the measured mass differs by more than 10^{24} . Explaining this hierarchy would require enormous fine-tuning cancellations [10, 11].

This has motivated theories proposing new physics at the TeV scale that stabilize the electroweak scale (ex: supersymmetric particles) and avoid this "unnatural" fine-tuning [12]. Several BSM theories such as supersymmetry or compositeness introduce mechanisms or symmetries that cancel or dampen these large corrections using partner particles around the weak scale [13]. However, the LHC has yet to discover particles or phenomena supporting these theories [14].

1.1.2 *Dark Matter*

The Λ CDM model provides overwhelming evidence that non-luminous, non-baryonic dark matter comprises approximately 26% of the energy density of the Universe, whereas ordinary baryonic matter accounts for about 5% [15–17]. The SM has no viable dark matter candidate – particles like neutrinos are too light and feebly interacting to account for the observed dark matter [18].

Some BSM theories postulate new stable or long-lived particles for dark matter, such as the lightest supersymmetric particle [17, 19]. Several of these candidates could be measured at CERN; if dark matter is a weakly interacting massive particle (WIMP) at the electroweak scale, it could be produced in high-energy collisions [20]. Atlas and CMS have both performed searches for dark matter produced in proton-proton collisions, looking for missing transverse momentum signatures accompanying jets or leptons [21–23]. While these searches have not yielded a discovery, they have placed limits on the regions in which to search for dark matter [24, 25].

¹ Nearly 5 decades after its theoretical proposal [5].

1.1.3 Flavor Puzzles and Further Open Questions

The SM treats flavor as an input and does not explain why there are three generations or why their masses span many orders of magnitude [26–29]. Several theories addressing these flavor puzzles could be measured at the LHC [30–32].

Additionally, the SM cannot account for the baryon asymmetry of the universe [33, 34].

Lastly, neutrino oscillation experiments show that neutrinos have non-zero mass, but the SM assumes massless neutrinos [35, 36]. Explaining this mass requires BSM physics, such as heavy right-handed neutrinos, new symmetries, or additional Higgs fields [37, 38].

1.1.4 Searching for BSM Physics

The strong theoretical and experimental motivation for BSM phenomena necessitate searches for BSM physics. Traditional analyses at the LHC targeting specific models (supersymmetry, dark matter, etc.) have thus far been unsuccessful in finding evidence of BSM physics, leading many to model-agnostic searches. Moreover, model-agnostic searches may reveal BSM physics that has not yet been postulated. This paper explores model-agnostic searches to find BSM physics.

1.2 CHALLENGES IN JET-BASED SEARCHES

Many promising BSM scenarios (heavy resonances, supersymmetry, etc.) at the LHC involve high-energy quarks or final state gluons, which are measured experimentally in jets [39]. A jet is a collimated spray of particles (mostly hadrons) resulting from the hadronization of a quark or gluon produced in a high-energy collision. While many BSM models predict signatures with jets, the overwhelming rate of ordinary QCD jet production means that potential signals are typically buried in a massive background of SM processes.

Several challenges arise model-agnostic jet-based anomaly detection:

1. High dimensionality
2. Noise and uncertainty
3. Interpretability

HIGH DIMENSIONALITY Jets are high-dimensional objects. Unlike a lepton or photon which can be described by a four-momentum, a jet can contain dozens of particles [40]. Each constituent particle has its own kinematic properties (momentum, angle, particle type) that helps form a complex radiation pattern. Important information about the jet’s origin is encoded in its substructure: for instance, a jet originating from the decay of a heavy particle (like a boosted W boson) usually has two or three distinct energy clusters (prongs) within it, whereas a typical quark- or gluon-initiated jet from QCD will have a more uniform, single-prong energy flow [41]. Experimentally, one uses a set of high-level observables to characterize jet substructure (such as mass, τ_{21} N -subjettiness, energy correlation functions), but any fixed set of observables may not capture the full

complexity of the data [40]. With such high dimensionality, one must consider how to balance reasonable compute (e.g. reasonably sized model inputs) and the loss of essential information.

NOISE AND UNCERTAINTY Detector effects and environmental noise complicate jet analyses. The CMS detector has finite resolution and acceptance [42]. Jets are reconstructed from detector signals where imperfections in reconstruction, calibration uncertainties, and pileup all affect the measured jet properties [43].

INTERPRETABILITY As with most "black box" models, interpretability remains a concern. A deep-learning-based anomaly detection approach often yields an abstract "anomaly score." Understanding what physical characteristics make an event anomalous can be challenging. Recent approaches attempt to improve interpretability by correlating anomaly scores with known observables.

1.3 RELATED WORK

Unsupervised anomaly detection methods identify signals without examples of specific new physics signals while training. Unlike supervised learning, where classifiers are trained on signal versus background samples for a particular theory, unsupervised methods learn the structure of known background and then flag outliers. Various unsupervised anomaly detection techniques have been proposed and are being explored by the CERN community [44].

AUTOENCODERS AND VARIATIONAL AUTOENCODERS Early applications of autoencoders (AEs) demonstrated their use for model-agnostic anomaly detection by identifying poorly reconstructed signals [45], although the model suffered from sparsity and failure to generalize [46]. Subsequent work evaluated multiple AE variants for real-time or web-based processing [47–49]. The probabilistic latent space of Variational Autoencoders (VAEs) enabled more robust outlier scoring [50, 51]. Further variations incorporated normalizing flows into the VAE latent space to capture complex feature distributions [52].

GENERATIVE AND DIFFUSION MODELS Generative Adversarial Networks (GANs) have been combined with autoencoders to create models such as GAN-AE [53]. Diffusion models have been used to construct background templates for analysis, demonstrating robustness to statistical fluctuations [54]. Normalizing flows have also been employed [55] with extensions such as surjective flows [56].

WEAKLY SUPERVISED LEARNING Weak supervision methods such as Classification Without Labels (CWoLa) and Mass Unspecific Supervised Tagging (MUST) enable training classifiers using mixed signal-background samples or control regions [57, 58]. These strategies avoid the need for precise simulation and allow data-driven classifier training. CWoLa has shown to match or exceed fully supervised models in real-data scenarios, while MUST enables resonance searches without biasing mass spectra.

CLUSTERING AND OUTLIER DETECTION Traditional clustering algorithms have also been applied to LHC data. Methods such as self-organizing maps (SOMs) and k-means

enable unsupervised anomaly localization [59]. These approaches offer increased interpretability over the more "black box" models.

QUANTUM MACHINE LEARNING Quantum machine learning (QML) offers novel architectures for anomaly detection. Quantum autoencoders, hybrid classical–quantum classifiers, and quantum similarity learning models have been proposed [60–63].

GRAPH NEURAL NETWORKS EdgeConv-based GNNs have been used within autoencoder frameworks to reconstruct graph-structured data and isolate deviations from QCD-like topologies [60, 64]. GNNs can capture both local and global event features, making them a strong candidate for structured anomaly detection.

OTHER APPROACHES Additional approaches include likelihood-free inference, permutation-invariant networks, tree-based models, and hybrid anomaly metrics [65, 66]. Additionally, contrastive learning approaches train models to differentiate between similar and dissimilar event pairs in a self-supervised manner. These techniques have been used to generate low-dimensional embeddings that preserve anomalous structure and suppress background variance [67].

1.3.1 Benchmarking Datasets

Benchmarking efforts like the LHC Olympics 2020 [68] and Dark Machines Challenge [69] have evaluated many of these techniques, demonstrating their performance across a range of simulated signal types. That being said, benchmarking datasets for HEP data are often limited and narrow.

THESIS ROADMAP

This thesis investigates unsupervised and model-agnostic anomaly detection strategies for jets at the LHC, with a focus on graph-based deep learning architectures. The structure of the thesis is as follows:

- Chapter 1: Introduction
Introduces the limitations of the Standard Model and motivates the search for Beyond the Standard Model (BSM) physics. Outlines the rationale for anomaly detection approaches and jet-based searches at the LHC.
- Chapter 2: Physics and Jet Representations
Provides the necessary physics background, including the structure of LHC detectors, jet formation and substructure, and key observables for identifying different jet types. Discusses the features used to represent jets and the distinctions between QCD and W jets.
- Chapter 3: Machine Learning for Jet Anomaly Detection
Reviews prior work in anomaly detection with jets, categorizing models by architecture (autoencoders, VAEs, graph-based methods, transformers). Introduces the relevant machine learning concepts, such as message passing, EdgeConv, and reconstruction-based anomaly scores

- **Chapter 4: Methods**

Describes the simulated dataset, the preprocessing pipeline for charged particles, and the construction of features and graphs. Defines the architectures used (autoencoder, transformer, graph autoencoder, and classifier), their training procedures, and the evaluation setup.

- **Chapter 5: Results and Interpretation**

Presents ROC curves and AUC metrics for all models. Analyzes the relative performance of each architecture and discusses the implications of weak separation between signal and background.

- **Chapter 6: Discussion and Future Work**

Summarizes the findings and their implications for model-agnostic searches at the LHC. Proposes several directions for future research, including more discriminative features, alternate graph strategies, real-data validation, and improved interpretability tools.

2.0.1 *Quantum Chromodynamics (QCD) Jets*

In a high-energy pp collision, when a quark or gluon is produced with large transverse momentum, it cannot exist as a free colored particle due to color confinement. Instead, the parton undergoes a parton shower: it radiates gluons and sometimes split into quark–antiquark pairs, which may also radiate. This showering continues until the partons reach low energies, when they hadronize. The resulting spray of hadrons is observed in the detector as a jet: a collimated cluster of particle tracks and calorimeter energy deposits. QCD substructure often includes one cluster of high p_T particles in the center of the jet, although jet substructure is highly varied due to the multitude of processes by which QCD jets are generated.

QCD jet production is the dominant high- p_T process at the LHC. Most commonly, QCD jets are produced in $2 \rightarrow 2$ parton scatterings. The main channels for jet production are:

1. quark–quark (qq) scattering,
2. quark–gluon (qg) scattering,
3. gluon–gluon (gg) scattering.

Based on the parton distribution functions, at LHC energies, gluon processes are significant, making gg and qg scatterings a large component of jet production. Quark-initiated jets become more prevalent at high momentum fractions.

Although quark and gluon jets have different characteristics, these jets are treated as the same for the purpose of this analysis. We use QCD jets as our background and W boson jets as our "signal-like" jets.

2.0.2 *W Boson Jets ($W+jets$)*

The W boson is a heavy gauge boson in the SM (with mass ~ 80.4 GeV) that often decays into hadrons or leptons. Here, we consider $W+jets$ as a proxy for a BSM signal within an anomaly detection framework. The idea is that decays of the W boson can be treated as a known "anomaly" – if our algorithms can detect W jets as deviations from QCD, they may be able to detect unknown signals of similar or larger differences.

A W boson decays about 67% of the time to quark-antiquark pairs, producing two jets. If the W is produced with low momentum, these two quarks hadronize into two well-separated jets, so the event contains two distinct jets. If the W is produced with high transverse momentum, its decay products are emitted close by, and the two resulting quark jets can merge into a single fat jet in the detector. Such a jet will have a mass around m_W and an internal two-prong substructure (two hard cores corresponding to the two quarks) 2.1. This is a very characteristic signature: a jet with two subjets and a consistent total mass.

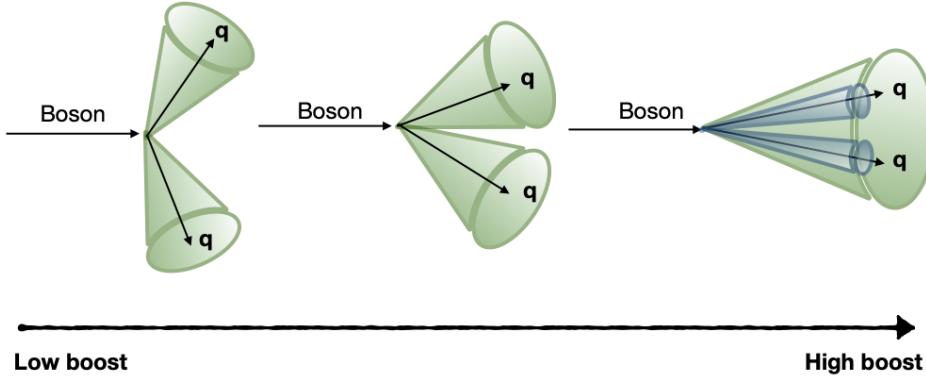


Figure 2.1: Jets arising from a W boson decaying to two quarks with increasing transverse momentum. Highly boosted bosons result in a jet with two substructures corresponding to the two original quarks. Lower transverse momentum jets result in two distinct jets. Image credit: [70].

2.0.3 Measuring Jets: The Detector

The CMS experiment is a detector at the LHC designed to measure particles emerging from high-energy $p p$ collisions [72, 73]. The detector consists of layered subsystems for tracking, calorimetry, and muon detection. Below, we summarize the key components relevant to measurements and the reconstruction chain that transforms raw signals into physics objects.

TRACKING SYSTEM Closest to the interaction point is a silicon-based tracker. It reconstructs charged particle trajectories, yielding their transverse momentum p_T , direction, and impact parameters. Momentum resolution is typically 1–10% for $p_T \sim 10\text{--}100 \text{ GeV}$ [73].

CALORIMETRY Surrounding the tracker are calorimeters that measure particle energy. The Electromagnetic Calorimeter (ECAL) measures electrons and photons via electromagnetic cascades. CMS uses high-granularity lead tungstate crystals with $\sim 1\%$ energy resolution for high-energy electrons and photons. The Hadronic Calorimeter (HCAL) captures energy from hadrons through nuclear interactions with a coarser resolution (5–10%). Combined ECAL and HCAL responses yield jet energy and contribute to missing transverse energy estimates.

MUON SYSTEM Muons are measured in dedicated outer chambers. CMS uses its magnet’s return yoke. These systems provide momentum resolution of 5–10% for muons up to several hundred GeV, enabling identification of high- p_T muons.

TRIGGER AND DATA ACQUISITION Given the 40 MHz collision rate, CMS implements multi-level trigger systems. A hardware-based Level-1 trigger uses calorimeter and muon information to reduce the rate to $\sim 100 \text{ kHz}$. The High-Level Trigger (HLT) then applies refined algorithms to select events for permanent storage at a few kHz.

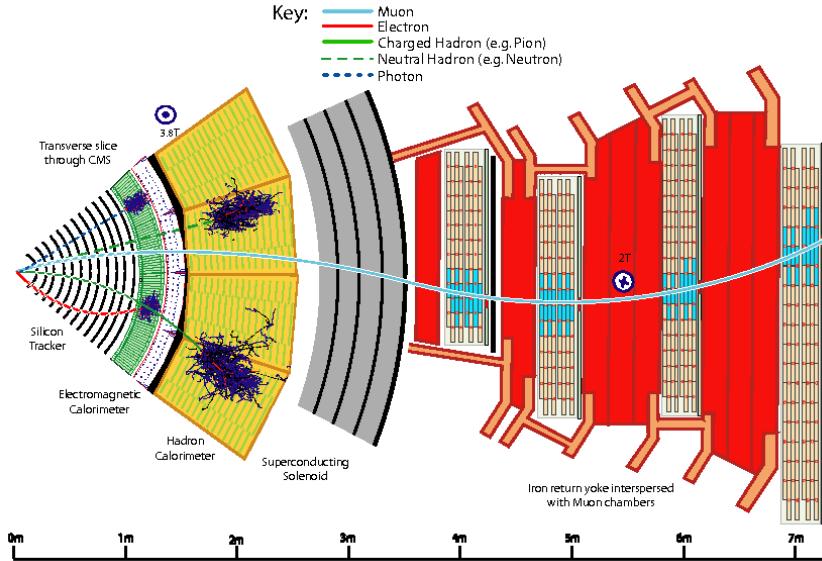


Figure 2.2: Schematic of CMS. Starting from the interaction point at the center, the main subdetectors are: the inner tracking system (silicon pixel and strip detector) in a solenoidal magnetic field that curves charged particle trajectories; the electromagnetic calorimeter to measure energy of electrons and photons via electromagnetic showers; the hadronic calorimeter to absorb and measure energy of hadrons; and the outer muon spectrometer to identify muons. Layers of the detector are optimized to measure properties of different particle types, allowing complete reconstruction of an event. Image credit: [71]

EVENT RECONSTRUCTION Offline processing reconstructs high-level objects from detector signals:

- Vertexing: Primary and pile-up interaction vertices are identified along the beam axis.
- Track Fitting: Charged-particle tracks are reconstructed from hits in the tracker.
- Particle Identification: Tracks and calorimeter clusters are combined to classify particles as electrons, muons, hadrons, or photons.
- Jet Clustering: Identified particles are clustered into jets, typically with radius $R = 0.4$ or 0.6 . Jet energy corrections and b -tagging algorithms are then applied.
- Lepton/Photon Reconstruction: Electrons and photons are identified by ECAL signatures with or without matched tracks; muons are matched across tracker and muon chambers.
- Missing Transverse Momentum: E_T is computed as the negative sum of reconstructed particle momenta.

These observables can feed anomaly detection algorithms directly in end-to-end ML models.

3.0.1 Autoencoders and Variants

Autoencoders are unsupervised neural networks designed to learn an efficient compression of input data and then reconstruct the original input from this compressed representation. An autoencoder consists of an encoder function f_θ that maps an input x (e.g. a representation of a jet) to a latent vector z in a lower-dimensional latent space, and a decoder function g_ϕ that maps z back to a reconstructed input \hat{x} . We denote this process as $z = f_\theta(x)$ and $\hat{x} = g_\phi(z)$ [74, 75]. The networks f_θ and g_ϕ are trained jointly by minimizing a reconstruction loss L_{rec} that measures the difference between the original and reconstructed inputs. We choose to use the mean squared error (MSE) between x and \hat{x} :

$$L_{\text{rec}}(x) = \|x - g_\phi(f_\theta(x))\|^2$$

which the autoencoder seeks to minimize for the training data [76].

For anomaly detection, we train the autoencoder on a sample of QCD jet events and then evaluate its reconstruction error on new data (W boson jets). Jets that are out-of-distribution are expected to yield larger reconstruction errors L_{rec} because the autoencoder cannot reproduce features it has not learned.

3.0.2 Convolutional Autoencoders

Convolutional autoencoders specialize the encoder and decoder as convolutional neural networks (CNNs), suitable for image-like representations of jets [77, 78]. The encoder uses learned filters to extract spatial features and the decoder reconstructs the image from latent features.

A single convolutional layer computes:

$$(X * W)(u, v) = \sum_{i,j} W(i, j) X(u + i, v + j)$$

with the loss computed over pixels as:

$$L_{\text{rec}} = \sum_{u,v} \|X(u, v) - \hat{X}(u, v)\|^2$$

3.0.3 Variational Autoencoders

Variational autoencoders (VAEs) extend autoencoders by introducing a probabilistic latent space with a prior $p(z)$ and approximating posterior $q_\phi(z|x)$ [79, 80]. The loss combines a reconstruction term with Kullback–Leibler (KL) divergence regularization:

$$L_{\text{VAE}}(x) = -\mathbb{E}_{z \sim q_\phi(z|x)} [\log p_\theta(x|z)] + \text{KL}(q_\phi(z|x) \| p(z))$$

The KL term enforces structured latent representations and suppresses overfitting. Variants such as β -VAEs further control the trade-off between compression and regularization.

3.0.4 Transformer-Based Models

Transformers apply self-attention to variable-length inputs, making them suitable for jets as unordered particle sets [81]. Each input token (particle) is encoded as:

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

where Q , K , and V are linear projections of the input features. Permutation invariance is achieved by omitting positional encodings and symmetrically pooling outputs. Transformers detect complex global anomalies and are highly expressive but computationally intensive.

3.0.5 Graph Autoencoders with EdgeConv

Jet events can be modeled as graphs with particles as nodes and edges encoding spatial or learned proximity. In graph neural networks, information is propagated through the graph by message passing, where each node aggregates information from its neighbors to update its own representation. This allows the network to learn relationships between particles. A common implementation is EdgeConv,

$$x'_i = \max_{j \in \mathcal{N}(i)} h_\Theta(x_i \parallel (x_j - x_i)) ,$$

where h_Θ is a neural network and \parallel denotes concatenation [82]. Graph autoencoders use EdgeConv layers for encoding and decoding. The loss includes node feature and (optionally) edge reconstruction. Graph architectures are robust to variable jet size and capture substructure but require careful design of graphs.

4.1 DATA SAMPLE

Datasets were simulated via MadGraph5_aMC@NLO [83], Pythia8 [84], Geant4 [85], and NanoAOD for QCD, W+ jets, and Hbb jets. For an overview of the simulation software and the workflow, see Table 4.1 and Figure 4.1.

4.1.1 Data Flow

MadGraph5_aMC@NLO was used to generate the hard scattering process at 13 TeV with MadGraph-MLM used for parton-jet matching interfaces.

Pythia8 was subsequently used to generate parton showering and hadronization based on our current understanding of the Standard Model. TuneCP5 parameter tuning was used. Data is consistent with the 2015-2018 CMS Run 2.

The generated events were passed to Geant4, which simulates how these particles would interact with the CMS detector, producing simulated raw detector signals. The simulated signals are processed using CMSSW to reconstruct high-level physics objects (electrons, muons, jets, etc.). These reconstructed objects are stored in the NanoAOD format.

The datasets were provided via ROOT format, an object-oriented framework designed for large-scale data analysis and storage [86]. Datasets were generated for QCD jets as the background events and W events as the signal.

4.1.2 Event Selection and Cutoff Criteria

The primary data attributes of interest are described in Table 4.2.

We apply selection cuts and cutoff criteria to each event based on several of the event's FatJet attributes. We apply a lepton veto such that each FatJet must be sufficiently separate in the eta-phi plane from any reconstructed leptons (electrons, taus, muons); this helps reduce background from processes that produce leptons. Any lepton with $p_T > 20.0$ is required to have a distance $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} > 0.4$.

We implement generation-matching to select FatJets that originate from hard-scattering processes of interest. Each FatJet must be matched to a generator-level jet within a 0.4 radius in the eta-phi plane; this ensures that each FatJet corresponds to a parton-level object produced in the primary interaction.

Kinematic cuts are applied to the FatJets to ensure they have sufficient energy; each FatJet must have a total $p_T > 200.0$ (transverse momentum) and a total $|\eta| < 2.1$ (absolute pseudorapidity).

After the above cuts, events with one or more FatJet are kept. If there is more than one FatJet per event, the one with the highest transverse momentum is selected.

For each FatJet, the associated particle flow candidates (PFCands) are identified; each PFCand represents a fundamental particle resulting from jet fragmentation that can be used to extract jet substructure for future analysis. PFCands are selected if they

Table 4.1: Software for CMS Data Generation and Processing

| Software | Purpose | Role in Simulation |
|----------------------|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| MadGraph5_aMC@NLO | Matrix Element Generator | Simulates hard scattering based on theoretical calculations based in the SM. Outputs events often in the Les Houches Event (LHE) format. |
| Pythia8 | Parton Shower and Hadronization | Simulates the parton showering and hadronization. |
| Geant4 | Detector Simulation | Simulates the passage of particles through the components of the CMS detector, produces simulated raw detector signals. |
| CMSSW (CMS Software) | Reconstruction Framework | Reconstructs high-level physics objects (e.g., electrons, muons, jets, photons) as they would be observed in the detector. |
| NanoAOD | Data Format | A compressed and streamlined data format for storing reconstructed physics objects and their properties. |
| ROOT | Data Format | An object-oriented framework for large-scale data analysis and storage. |

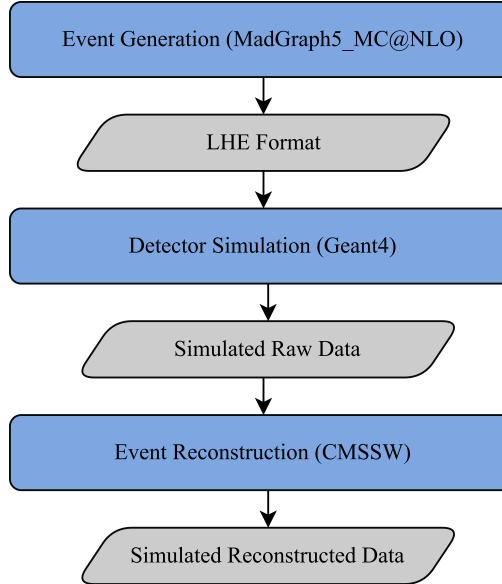


Figure 4.1: An overview of the simulation workflow. Events are generated, then detector signals are simulated. Those simulated signals are reconstructed into the original events, from which we get the data we use in our analysis.

are within certain bounds relative to their parental FatJet: $|\eta_{FatJet} - \eta_{PFCand}| < 0.1$, $|\phi_{FatJet} - \phi_{PFCand}| < 0.1$, and $|p_{T,FatJet} - p_{T,PFCand}| < 1$.

For each PFCand in the FatJet, several fields are stored. The only jet-level information stored is the event's LHE, which is used for energy normalization later.

4.2 DATA CLEANING AND FEATURE ENGINEERING

4.2.1 Jet Level

The p_T of each PFCand is log-normalized by the FatJet p_T :

$$p_{T,PFCand} = \log\left(\frac{p_{T,PFCand}}{p_{T,FatJet}}\right)$$

The log transformation reduces the skewness of the p_T distribution, making the features more Gaussian-like. Eta and phi are also normalized:

$$\eta_{PFCand} = \eta_{PFCand} - \eta_{FatJet}$$

$$\phi_{PFCand} = \phi_{PFCand} - \phi_{FatJet}$$

This ensures that each FatJet selected has values within the same ranges and that all the values are within the same reference frames. This normalization procedure helps to remove the dependence on the overall momentum and direction of the FatJet, allowing the model to focus on the internal structure and relationships between the constituent particles.

For now, we only consider charged particles in our analysis to avoid using sparse matrices (lots of uncharged particles have errors or 0.0 values for certain attributes, making the distributions wonky). This keeps our particles restricted to particles with PDG IDs of -11, 11, -13, 13, -211, and 211.

To capture the particle type information of the charged particles within the FatJet, we one-hot-encode the pdgId of each PFCand.

For each PFCand within a FatJet, we calculate the ratio of the impact parameter in the z-direction (dz) to its uncertainty ($dzErr$). The absolute value of this ratio, $|\frac{dz}{dzErr}|$, is capped at 5 to remove outliers and potentially mis-simulated tracks. This cut also helps select tracks originating from the primary vertex of the event. Similarly, we perform the same procedure for the impact parameter in the xy-plane ($d0$) and its uncertainty ($d0Err$), again capping the absolute value of $|\frac{d0}{d0Err}|$ at 5.

4.2.2 Dataset Level

We apply a per-feature scaling to the entire dataset. For each continuous attribute, we determine the 16th and 84th percentiles of its distribution across all training samples. These percentiles, p_{-16} and p_{+84} respectively (corresponding to approximately ± 1 standard deviation for a Gaussian distribution), provide estimates of the distribution's spread while remaining less sensitive to extreme outliers (like the standard deviation). We then scale each individual sample x using the formula:

$$\frac{x - p_{-16}}{p_{+84} - p_{-16}} \times 2 - 1$$

This scaling transforms each feature to have a distribution centered around 0 with a spread roughly between -1 and 1, normalizing the features to a common scale. Note that one-hot-encoded variables are kept in their original binary format.

[data scaling/normalization handling missing values figure: features before and after preprocessing](#)

4.3 DATA STRUCTURE

4.3.1 Array Format

When input into an autoencoder, each event is encoded as a multi-dimensional array. For example, if the features are η , ϕ , and p_T , the array dimension in x is η , the array dimension in y is ϕ , and the "color" of the array is p_T . This scales with more and more features. To perfectly separate all the particles, we would need an incredibly large array. However, that size of an array will not run on our computational resources. Thus, we introduce binning wherein we take the particle in each bin that has the highest p_T and we choose a total array size of 32×32 . Note that most of the bins carry no information, as the matrix is highly sparse. For an example of the array format on jet data, see Figure [4.2](#).

4.3.2 Graph Format

Each graph in our graph-based representation is a different event. The fundamental unit of our graph representation is a node, which corresponds to a particle (PFCand). For each reconstructed particle i in an event, a node v_i is created. The edges in the graph connect the nodes during the message-passing operations of the autoencoder; thus, the edges represent the relationships and interactions between the reconstructed particles.

Table 4.2: Particle Attributes and Their Information

| Attribute Name | Description | Units | Preprocessing Steps |
|-------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Pseudorapidity (η) | Spatial coordinate describing the angle of the particle relative to the beam axis. | N/A | Normalized by subtracting the FatJet η : $\eta_{PFCand} - \eta_{FatJet}$ where $ \eta_{PFCand} - \eta_{FatJet} < 0.1$, else ignored. |
| Azimuthal Angle (ϕ) | The angle of the particle around the beam axis. | radians | Normalized by subtracting the FatJet ϕ : $\phi_{PFCand} - \phi_{FatJet}$ where $ \phi_{PFCand} - \phi_{FatJet} < 0.1$, else ignored. |
| $\frac{dz}{dzErr}$ | The signed longitudinal impact parameter (dz) of the particle with respect to the primary vertex, divided by its uncertainty ($dzErr$). | N/A | Absolute value taken: $ \frac{dz}{dzErr} $, then capped at 5. Distribution across the whole dataset scaled to mean 0 and 16th and 84th percentiles at ± 1 . |
| $\frac{d_0}{d_0Err}$ | The signed transverse impact parameter (d_0) of the particle with respect to the primary vertex, divided by its uncertainty (d_0Err). | N/A | Absolute value taken: $ \frac{d_0}{d_0Err} $, then capped at 5. Distribution across the whole dataset scaled to mean 0 and 16th and 84th percentiles at ± 1 . |
| PDG ID (One-Hot Encoded) | The Particle Data Group identifier, uniquely identifying the particle type. This is represented using one-hot encoding, resulting in multiple binary features (one for each considered particle type). | N/A | One-hot encoded for charged particles only. Neutral particles ignored. |
| Transverse Momentum (p_T) | The component of the particle's momentum perpendicular to the beam axis. | GeV | Log-normalized by the FatJet p_T : $\log\left(\frac{p_{T,PFCand}}{p_{T,FatJet}}\right)$. Distribution across the whole dataset scaled to mean 0 and 16th and 84th percentiles at ± 1 . |

Table 4.3: The attributes and preprocessing steps for each particle in the analysis.

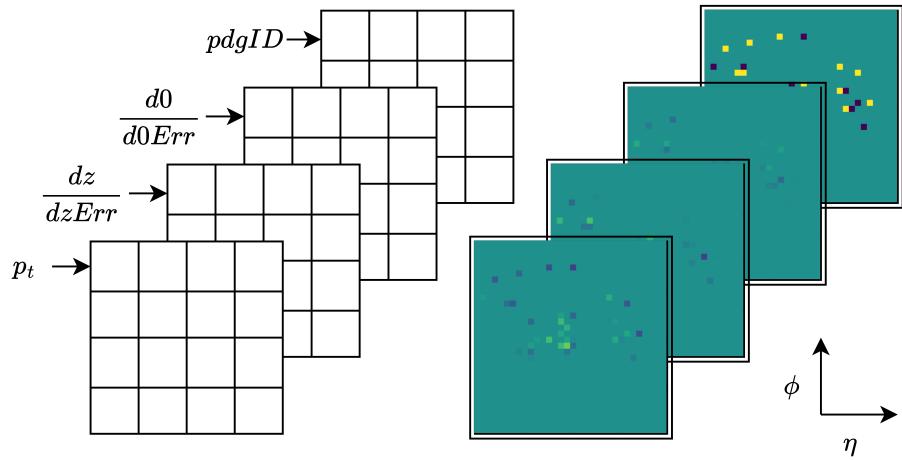


Figure 4.2: Example array structure of a jet. Array features are binned in eta-phi space, with the color channel representing the property of interest. Array jets are highly sparse and may lose information depending on the binning size and strategy.

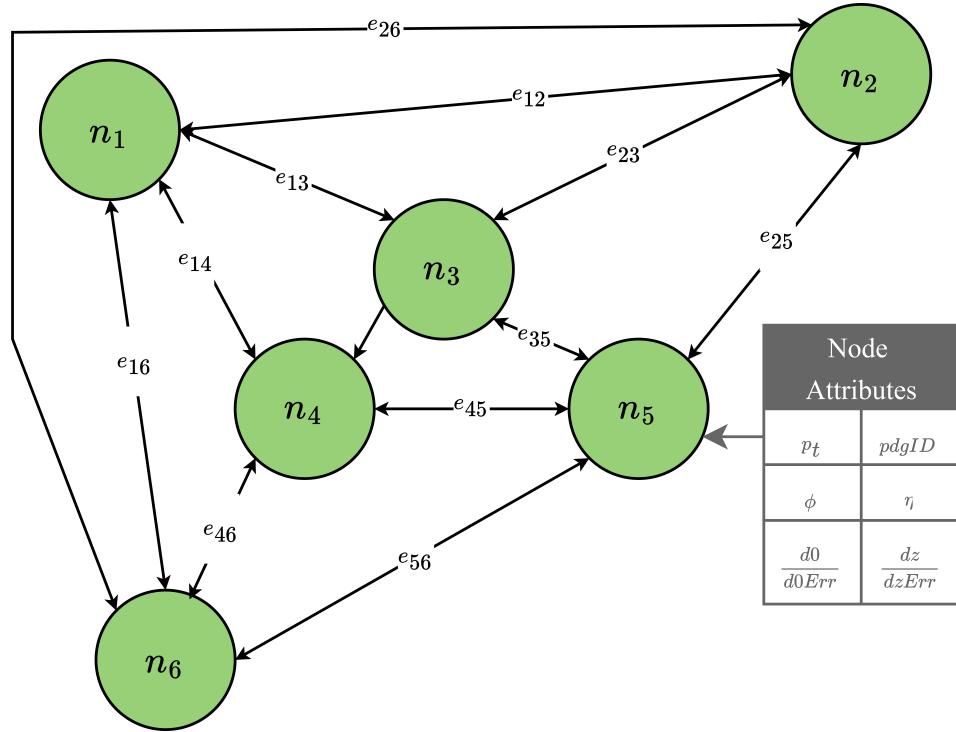


Figure 4.3: Example Graph structure with K-nearest-neighbors connecting the 4 nearest nodes in the feature space. Note that the edges are exaggerated here, and do not have a length(weight).

We construct proximity-Based edges. Edges are calculated based on their proximity in the η, ϕ plane. An edge is established between two particles i and j if their separation in the $\eta - \phi$ plane is within the closest K neighbors for that particle. The alternative approach is to define $\Delta R < \text{Delta } R_{\max}$, although this approach was not chosen because it would build a local connectivity around each particle and result in a variable number of edges for each node.

Note that all of these edges are non-directed because the information is equal between each particle.

Once the graph structure is defined, each node and edge is associated with a set of features. Node features include $\log\left(\frac{p_{T,PFCand}}{p_{T,FatJet}}\right)$, $\frac{d0}{d0Err}$, $\frac{dz}{dzErr}$, one-hot-encoded pdgId, η , and ϕ . In a more advanced algorithm, the covariances of each element would also be included, although this information was not initially generated and thus had to be excluded.

Edge features, when utilized, provided information about the relationship between connected particles. For proximity-based edges, the feature could be ΔR_{ij} . For the feature-space-based edges, the feature could be the total Δ for all the features. Future work could look into categorical edges (same types of particles, similar p_T , one type of edge for those original KNN with a ΔR metric).

Graph formatting is implemented using Pytorch Geometric [87]; reference Figure 4.3 for a sample graph structure.

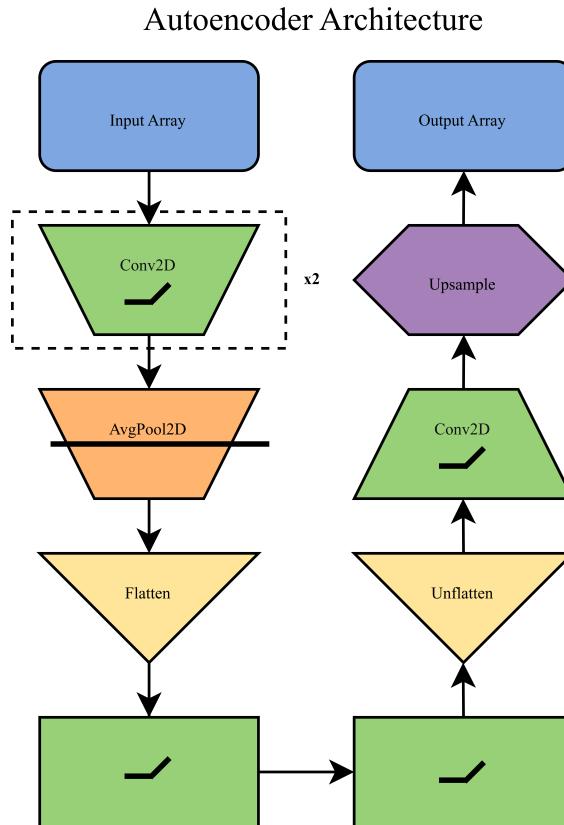


Figure 4.4: Autoencoder Architecture

4.4 MODEL IMPLEMENTATIONS

4.4.1 Autoencoder

Autoencoder architecture is intentionally minimal: a few fully connected layers with ReLU or tanh activations. The model is trained using a mean squared error (MSE) loss on node features only. Events that produce large reconstruction errors are flagged as anomalous. The simplicity of this model makes it a useful test for the information content of the chosen features and their ability to distinguish W jets from background. For a detailed overview of sample architecture, see 4.4.

4.4.2 Graph Autoencoder

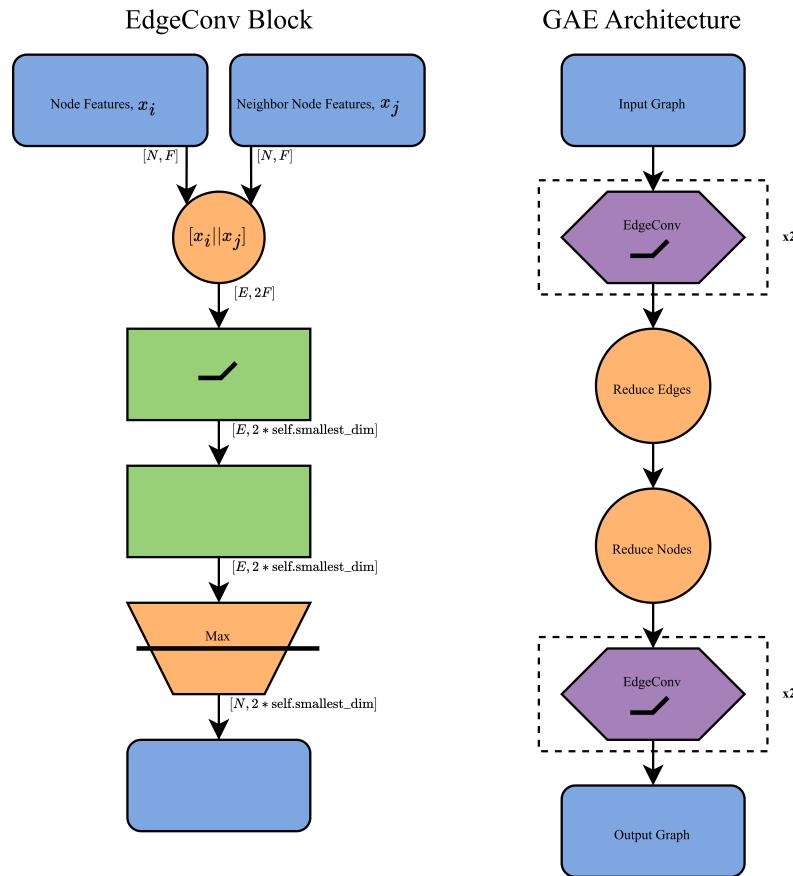


Figure 4.5: Graph Autoencoder Architecture

Graph autoencoder architecture is adapted from a highly simplified ParticleNet architecture, using EdgeConv blocks as its base. The encoder uses a series of EdgeConv blocks to propagate and embed node features, optionally reducing node count via pooling. The decoder mirrors this structure, reconstructing original node features from the latent graph. This design allows the model to capture local substructure and relational information between particles, which is crucial for identifying jets with multi-prong decay patterns like those from boosted W bosons. Unlike the classifier, this model does

not use labels during training and flags out-of-distribution jets by their reconstruction loss. For a detailed overview of the architecture, see [4.5](#).

4.4.3 Baseline: Supervised Learning/Classification

As a baseline, we establish a classification model to compare against unsupervised models and determine an upper bound for unsupervised performance. This model shares the architecture of the graph autoencoder but replaces the decoder with a linear classification head. It uses the same EdgeConv layers to embed the graph-structured jet into a latent representation, followed by a final fully-connected layer that outputs class probabilities. While this supervised setup is not realistic for truly model-agnostic searches, it provides a useful upper bound for performance and helps contextualize the results of the unsupervised methods.

4.5 TRAINING, EVALUATION, AND TUNING

Unless otherwise specified, hyperparameters used are the following: the loss function was defined as the Mean Squared Error (`MSELoss()`); a batch size of 64 samples was utilized during training; the number of training epochs was fixed at 20; an initial learning rate of `1e5` was adopted for the optimization process; a 5-fold cross-validation strategy was implemented for robust model evaluation; and a weight decay penalty of `5e4` was applied to the model parameters to mitigate overfitting. The activation function used will be specified in the relevant sections describing the network architecture.

Models are evaluated using loss curves to monitor the presence of overfitting/underfitting, ROC curves, and anomaly score distributions.

5

RESULTS AND INTERPRETATION

5.1 CHARACTERIZING THE BACKGROUND MODEL

Each model is evaluated on its ability to distinguish signal-like (W jets) from QCD background using anomaly scores or classification logits. The ROC curve for each model is shown below.¹

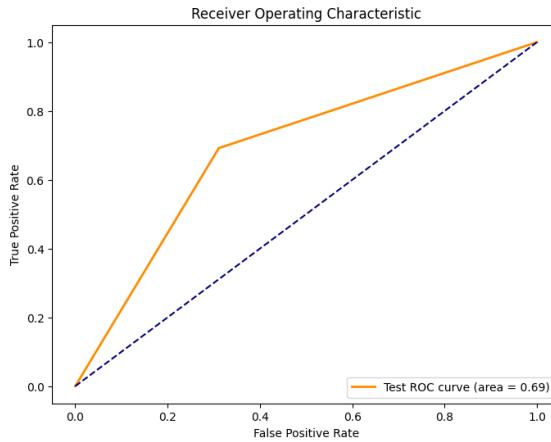


Figure 5.1: ROC curve for the supervised classification baseline. This model provides an upper bound for model performance, as it is trained directly to separate QCD from W jets. The observed AUC was 0.69, which establishes an upper bound on our other model performances with moderate separation between background and signal.

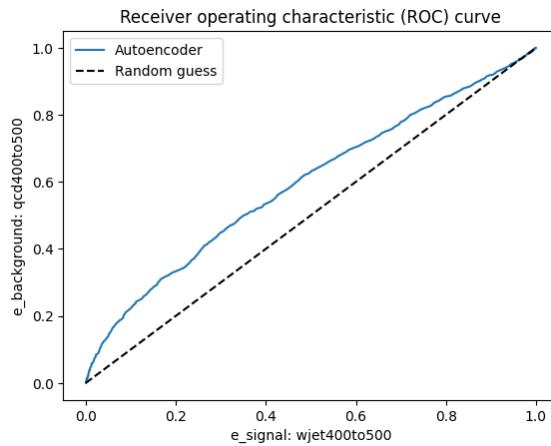


Figure 5.2: ROC curve for the standard autoencoder trained on QCD background. Signal-like jets produce higher reconstruction error. AUC of 0.594, showing very slight separation.

¹ Earlier versions of the autoencoder and graph autoencoder models contained a bug in the evaluation script that led to inflated performance metrics, which have since been corrected. Thus, the curves shown here may be worse performing than expected by the reader.

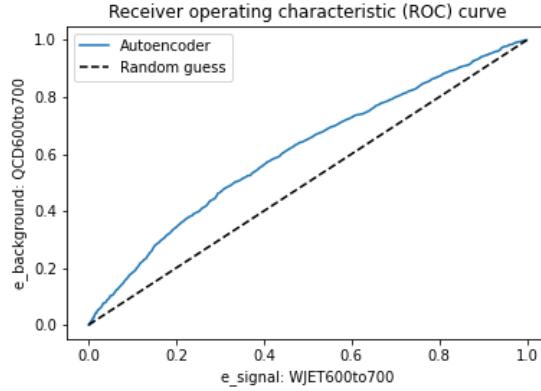


Figure 5.3: ROC curve for the graph autoencoder using EdgeConv layers. The GAE shows improved similar sensitivity over the standard AE with an AUC of 0.606.

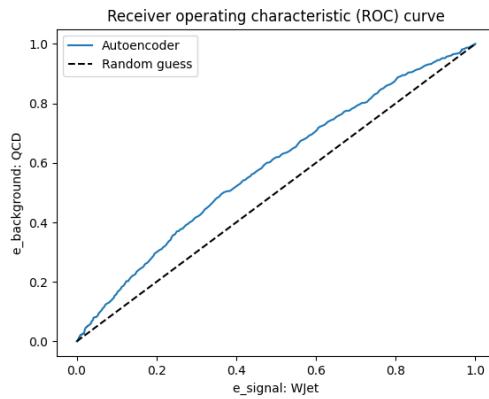


Figure 5.4: ROC curve for the Transformer model operating on variable-length particle sets. Its performance depends on capturing global structure via self-attention. The performance is comparable to the other models, with an AUC of 0.584.

5.2 INTERPRETATION AND COMPARISON

Table 5.1 summarizes the performance of each model. The supervised classifier, trained directly on labeled QCD and W jets, achieved an AUC of 0.69, which sets an empirical upper bound for the performance of unsupervised methods. This moderate separation indicates that even with labels, distinguishing between W jets and QCD jets is nontrivial. Examples of autoencoder reconstructions with high loss and low loss are shown in Figures 5.5 and 5.6, showing that the autoencoder best reconstructs jets with a single high-momenta cluster of particles.

Unsupervised models show only marginal separation between W jets and QCD. The standard autoencoder yields an AUC of 0.594, while the graph autoencoder slightly improves this to 0.606. The transformer model performs comparably, with an AUC of 0.584. These results suggest that, under the current configuration and data representation, the W jets are not strongly out-of-distribution relative to the QCD background. That is, the anomaly detection task is inherently difficult on this dataset.

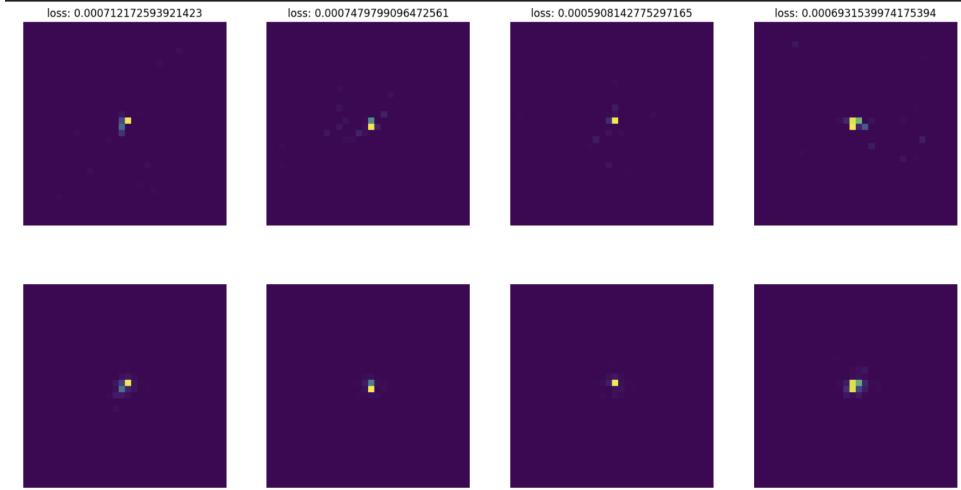


Figure 5.5: Signal events with the lowest reconstruction loss from the autoencoder. Original input arrays are shown on the top row, and corresponding reconstructions on the bottom row. These well-reconstructed events typically exhibit simple jet structures with a single, central high-momentum cluster.

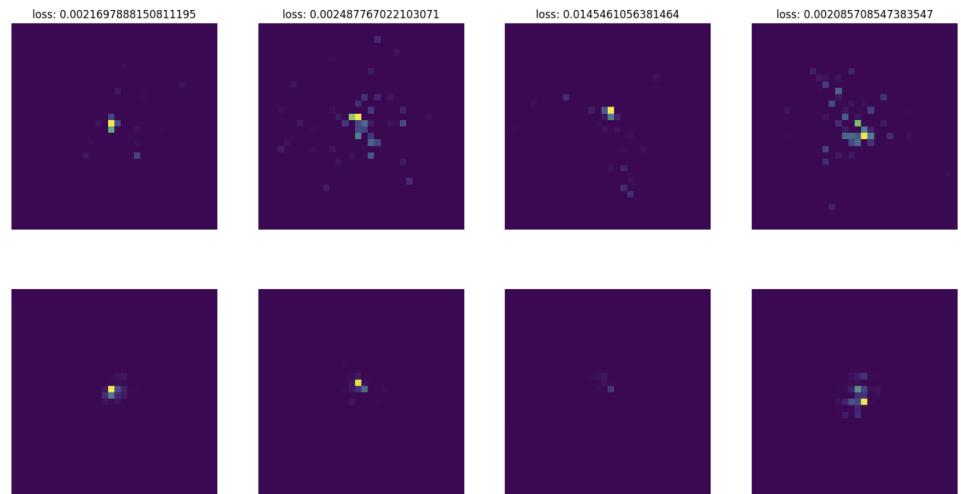


Figure 5.6: Signal events with the highest reconstruction loss from the autoencoder. Original input arrays are shown on the top row, and corresponding reconstructions on the bottom row. Poorly reconstructed events often contain multiple particle clusters, particles outside of the central high- p_T cluster, or off-center high-momentum regions.

Table 5.1: AUC scores and training parameters for each model. Hyperparameters are kept consistent where possible for fair comparison.

| Model | AUC | Learning Rate | Batch Size | Weight Decay | # Signal | # Background |
|----------------|------------|----------------------|-------------------|---------------------|-----------------|---------------------|
| Classification | 0.690 | 1×10^{-3} | 64 | 0 | 50,000 | 50,000 |
| AE | 0.594 | 1×10^{-5} | 64 | 5×10^{-4} | 70,000 | 120,000 |
| GAE | 0.606 | 1×10^{-5} | 64 | 5×10^{-4} | 70,000 | 120,000 |
| Transformer | 0.584 | 1×10^{-5} | 64 | 5×10^{-4} | 70,000 | 120,000 |

6

DISCUSSION AND FUTURE WORK

6.1 INTERPRETATION OF RESULTS

The results presented in Chapter 5 demonstrate that model-agnostic anomaly detection using unsupervised learning on jet substructure remains a challenging task. The autoencoder, transformer, and graph-based methods yielded AUC scores in the range 0.58–0.61 when tasked with identifying W jets as anomalous within a QCD-dominated dataset. While these results show some sensitivity to signal-like jets, they fall short of the upper bound set by a supervised classifier trained with ground-truth labels.

The lack of significant performance difference across model architectures raises several important questions:

First, we consider whether the W jets are sufficiently distinct from the QCD background in this feature space. Although W jets possess characteristic substructure (such as two-prong decays), this signature may be blurred by detector effects or only make up a minority of the dataset. Moreover, since only charged particles are used, important neutral energy contributions are neglected, potentially limiting model performance.

Second, the models may be capturing small systematic differences between the signal and background datasets—such as differences in generator settings or detector simulation artifacts—rather than true physical differences in jet structure.¹

Ultimately, the differences in AUC are small enough that architecture choice may be less important than the choice of representation, feature engineering, or data domain. Importantly, because all models performed close to chance level, we did not pursue extensive hyperparameter tuning or advanced feature augmentation. With no architecture demonstrating promising baseline performance, further optimization efforts risk overfitting to noise or dataset-specific artifacts rather than uncovering genuine signal sensitivity. This bottleneck underscores the importance of validating separability in the input space before large-scale model development.

6.2 FUTURE WORK

Several directions are worth pursuing to build upon this work:

- Signal choice and benchmarking: W jets serve as a useful stand-in for structured anomalies but are not necessarily representative of the kinds of BSM signals unsupervised models are intended to detect. Future studies should incorporate well-motivated BSM signals with more distinct kinematics, such as long-lived particle decays, multi-prong jets, or semivisible jets. Datasets used should be standard in the field, such as LHC02020 or the Dark Machines Challenge.
- Feature space expansion: Including neutral particles, more features per particle, covariances, and global jet features may improve model sensitivity.

¹ This is a particular problem since there is no concrete information on how the datasets were generated.

- Advanced graph construction: This work used fixed- K nearest neighbor graphs in η - ϕ space. Learned edge definitions, edge weights, or dynamic neighborhood sizes could improve the ability to model physically relevant structure.
- Interpretability: Ongoing efforts are needed to connect latent representations or anomaly scores to interpretable physical features. Techniques such as saliency maps, latent traversals, or clustering in the latent space could help elucidate model behavior. Again, these were not explored due to the poor performance of the model.

BIBLIOGRAPHY

- [1] Sheldon L. Glashow. "Partial Symmetries of Weak Interactions." In: *Nucl. Phys.* 22 (1961), pp. 579–588.
- [2] Steven Weinberg. "A Model of Leptons." In: *Phys. Rev. Lett.* 19 (21 1967), pp. 1264–1266. DOI: [10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.
- [3] Abdus Salam. "Weak and Electromagnetic Interactions." In: (1968). Ed. by N. Svartholm.
- [4] Particle Data Group. "Review of Particle Physics." In: *Prog. Theor. Exp. Phys.* 2022.8 (2022), p. 083C01.
- [5] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons." In: *Phys. Rev. Lett.* 13 (16 1964), pp. 508–509. DOI: [10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [6] 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), pp. 1–29.
- [7] 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), pp. 30–61.
- [8] L. Susskind. "Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory." In: *Phys. Rev. D* 20 (1979), pp. 2619–2625.
- [9] E. Gildener and S. Weinberg. "Symmetry Breaking and Scalar Bosons." In: *Phys. Rev. D* 13 (1976), p. 3333.
- [10] R. Barbieri and G. F. Giudice. "Upper Bounds on Supersymmetric Particle Masses." In: *Nucl. Phys. B* 306 (1988), pp. 63–76.
- [11] G. 't Hooft. "Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking." In: *NATO Adv. Study Inst. Ser. B Phys.* 59 (1980), p. 135.
- [12] S. Dimopoulos and H. Georgi. "Softly Broken Supersymmetry and SU(5)." In: *Nucl. Phys. B* 193 (1981), pp. 150–162.
- [13] Stephen P. Martin. "A Supersymmetry Primer." In: *Adv. Ser. Direct. High Energy Phys.* 21 (2010), pp. 1–153. eprint: [hep-ph/9709356](https://arxiv.org/abs/hep-ph/9709356).
- [14] Nathaniel Craig. "The State of Supersymmetry after Run I of the LHC." In: *arXiv preprint* (2013). eprint: [1309.0528](https://arxiv.org/abs/1309.0528).
- [15] N. Aghanim et al. [Planck Collaboration]. "Planck 2018 results. VI. Cosmological parameters." In: *Astron. Astrophys.* 641 (2020), A6. eprint: [1807.06209](https://arxiv.org/abs/1807.06209).
- [16] N. Aghanim et al. "Planck 2018 results. Overview and legacy of the mission." In: *Astron. Astrophys.* 641 (2020), A1.
- [17] Dan Hooper Gianfranco Bertone and Joseph Silk. "Particle dark matter: Evidence, candidates and constraints." In: *Phys. Rep.* 405 (2005), pp. 279–390.

- [18] M. Kamionkowski G. Jungman and K. Griest. "Supersymmetric dark matter." In: *Phys. Rep.* 267 (1996), pp. 195–373.
- [19] Jonathan L. Feng. "Dark Matter Candidates from Particle Physics and Methods of Detection." In: *Ann. Rev. Astron. Astrophys.* 48 (2010), pp. 495–545.
- [20] Lars Bergström. "Non-baryonic dark matter: Observational evidence and detection methods." In: *Rept. Prog. Phys.* 63 (2000), pp. 793–841.
- [21] Daniel Abercrombie et al. "Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum." In: *Phys. Dark Univ.* 27 (2020), p. 100371.
- [22] M. Aaboud et al. [ATLAS Collaboration]. "Search for dark matter and other new phenomena in events with an energetic jet and large missing transverse momentum using the ATLAS detector." In: *JHEP* 01 (2018), p. 126.
- [23] A. M. Sirunyan et al. [CMS Collaboration]. "Search for new physics in the monophoton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV." In: *Phys. Lett. B* 779 (2018), pp. 283–306.
- [24] CMS Collaboration. "Search for dark matter produced in association with a leptonically decaying Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV." In: *JHEP* 10 (2022), p. 136.
- [25] ATLAS Collaboration. "Combination of searches for invisible Higgs boson decays with the ATLAS experiment." In: *Phys. Rev. Lett.* 122 (2021), p. 231801.
- [26] C.D. Froggatt and H.B. Nielsen. "Hierarchy of quark masses, Cabibbo angles and CP violation." In: *Nucl. Phys. B* 147 (1979), pp. 277–298.
- [27] Guido Altarelli and Ferruccio Feruglio. "Discrete flavor symmetries and models of neutrino mixing." In: *Rev. Mod. Phys.* 82 (2010), p. 2701.
- [28] K.S. Babu. "TASI Lectures on Flavor Physics." In: *arXiv preprint* (2009). eprint: [0910.2948](https://arxiv.org/abs/0910.2948).
- [29] G.C. Branco, L.avoura, and J.P. Silva. *CP Violation*. Oxford University Press, 2012.
- [30] D. Bailey S. Davidson and B. A. Campbell. "Model independent constraints on leptoquarks from rare processes." In: *Z. Phys. C* 61 (1994), pp. 613–644.
- [31] Martin Bauer and Matthias Neubert. "Minimal Leptoquark Explanation for the $R_{D^{(*)}}$, R_K , and $(g - 2)_\mu$ Anomalies." In: *Phys. Rev. Lett.* 116 (2016), p. 141802.
- [32] A.J. Buras and J. Girrbach. "Towards the Identification of New Physics through Quark Flavour Violating Processes." In: *Rept. Prog. Phys.* 77 (2014), p. 086201.
- [33] M. B. Gavela et al. "Standard Model CP-violation and Baryon Asymmetry." In: *Mod. Phys. Lett. A* 9 (1994), pp. 795–810.
- [34] Antonio Riotto and Mark Trodden. "Recent progress in baryogenesis." In: *Ann. Rev. Nucl. Part. Sci.* 49 (1999), pp. 35–75.
- [35] Y. Fukuda et al. [Super-Kamiokande Collaboration]. "Evidence for Oscillation of Atmospheric Neutrinos." In: *Phys. Rev. Lett.* 81 (1998), p. 1562.
- [36] Q.R. Ahmad et al. [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.

- [37] R.N. Mohapatra et al. "Theory of neutrinos: A White paper." In: *Rept. Prog. Phys.* 70 (2007), p. 1757.
- [38] Stephen F. King. "Models of Neutrino Mass, Mixing and CP Violation." In: *J. Phys. G* 42 (2015), p. 123001.
- [39] CMS Collaboration. "CMS highlights on searches for new physics in final states with jets." In: *arXiv preprint arXiv:2401.07172* (2024).
- [40] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. "Energy Correlation Functions for Jet Substructure." In: *Journal of High Energy Physics* 2013.06 (2013), p. 108. eprint: [arXiv:1305.0007](https://arxiv.org/abs/1305.0007).
- [41] CMS Collaboration. *Jet substructure – Physics Objects*. <https://cms-opendata-workshop.github.io/workshop2022-lesson-physics-objects/06-substructure/index.html>. 2022.
- [42] CMS Collaboration. *Performance of the pile up jet identification in CMS for Run 2*. <https://cds.cern.ch/record/2715906>. 2019.
- [43] CMS Collaboration. *How CMS weeds out particles that pile up*. <https://cms.cern/news/how-cms-weeds-out-particles-pile>. 2019.
- [44] ATLAS Collaboration. "Anomaly detection in ATLAS using unsupervised machine learning." In: *Eur. Phys. J. C* 81.12 (2021), p. 1110. DOI: [10.1140/epjc/s10052-021-09749-0](https://doi.org/10.1140/epjc/s10052-021-09749-0).
- [45] Marco Farina, Yuichiro Nakai, and David Shih. "Searching for New Physics with Deep Autoencoders." In: *Phys. Rev. D* 101 (2020), p. 075021. DOI: [10.1103/PhysRevD.101.075021](https://doi.org/10.1103/PhysRevD.101.075021).
- [46] Thorben Finke, Michael Krämer, Alessandro Morandini, Alexander Mück, and Ivan Oleksiyuk. "Autoencoders for unsupervised anomaly detection in high energy physics." In: *Journal of High Energy Physics* 2021.6 (June 2021). ISSN: 1029-8479. DOI: [10.1007/jhep06\(2021\)161](https://doi.org/10.1007/jhep06(2021)161). URL: [http://dx.doi.org/10.1007/JHEP06\(2021\)161](http://dx.doi.org/10.1007/JHEP06(2021)161).
- [47] Florencia Canelli, Annapaola de Cosa, Luc Le Pottier, Jeremi Niedziela, Kevin Pedro, and Maurizio Pierini. "Autoencoders for semivisible jet detection." In: *Journal of High Energy Physics* 2022.2 (Feb. 2022). ISSN: 1029-8479. DOI: [10.1007/jhep02\(2022\)074](https://doi.org/10.1007/jhep02(2022)074). URL: [http://dx.doi.org/10.1007/JHEP02\(2022\)074](http://dx.doi.org/10.1007/JHEP02(2022)074).
- [48] D. Abadjiev et al. "Autoencoder-Based Anomaly Detection System for Online Data Quality Monitoring of the CMS Electromagnetic Calorimeter." In: *Computing and Software for Big Science* 8.1 (June 2024). ISSN: 2510-2044. DOI: [10.1007/s41781-024-00118-z](https://doi.org/10.1007/s41781-024-00118-z). URL: <http://dx.doi.org/10.1007/s41781-024-00118-z>.
- [49] Sergei V. Chekanov, Wasikul Islam, Rui Zhang, and Nicholas Luongo. "ADFilter—A Web Tool for New Physics Searches with Autoencoder-Based Anomaly Detection Using Deep Unsupervised Neural Networks." In: *Information* 16.4 (Mar. 2025), p. 258. ISSN: 2078-2489. DOI: [10.3390/info16040258](https://doi.org/10.3390/info16040258). URL: <http://dx.doi.org/10.3390/info16040258>.
- [50] Taoli Cheng, Jean-François Arguin, Julien Leissner-Martin, Jacinthe Pilette, and Tobias Golling. "Variational autoencoders for anomalous jet tagging." In: *Physical Review D* 107.1 (Jan. 2023). ISSN: 2470-0029. DOI: [10.1103/physrevd.107.016002](https://doi.org/10.1103/physrevd.107.016002). URL: <http://dx.doi.org/10.1103/PhysRevD.107.016002>.

- [51] Ryan Liu, Abhijith Gandrakota, Jennifer Ngadiuba, Maria Spiropulu, and Jean-Roch Vlimant. *Fast Particle-based Anomaly Detection Algorithm with Variational Autoencoder*. 2023. arXiv: [2311.17162 \[hep-ex\]](https://arxiv.org/abs/2311.17162). URL: <https://arxiv.org/abs/2311.17162>.
- [52] Pratik Jawahar, Thea Arrestad, Nadezda Chernyavskaya, Maurizio Pierini, Kinga A. Wozniak, Jennifer Ngadiuba, Javier Duarte, and Steven Tsan. “Improving Variational Autoencoders for New Physics Detection at the LHC With Normalizing Flows.” In: *Frontiers in Big Data* 5 (Feb. 2022). ISSN: 2624-909X. doi: [10.3389/fdata.2022.803685](https://doi.org/10.3389/fdata.2022.803685). URL: <http://dx.doi.org/10.3389/fdata.2022.803685>.
- [53] Louis Vaslin, Vincent Barra, and Julien Donini. “GAN-AE: an anomaly detection algorithm for New Physics search in LHC data.” In: *The European Physical Journal C* 83.11 (Nov. 2023). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-023-12169-4](https://doi.org/10.1140/epjc/s10052-023-12169-4). URL: <http://dx.doi.org/10.1140/epjc/s10052-023-12169-4>.
- [54] Debajyoti Sengupta, Matthew Leigh, John Andrew Raine, Samuel Klein, and Tobias Golling. *Improving new physics searches with diffusion models for event observables and jet constituents*. 2023. arXiv: [2312.10130 \[physics.data-an\]](https://arxiv.org/abs/2312.10130). URL: <https://arxiv.org/abs/2312.10130>.
- [55] Debajyoti Sengupta, Samuel Klein, John Andrew Raine, and Tobias Golling. “CURTAINS Flows For Flows: Constructing Unobserved Regions with Maximum Likelihood Estimation.” In: *SciPost Phys.* 14.1 (2023), p. 001. arXiv: [2305.04646 \[hep-ph\]](https://arxiv.org/abs/2305.04646).
- [56] Rob Verheyen and et al. “Event Generation and Density Estimation with Surjective Normalizing Flows.” In: *SciPost Phys.* 13.3 (2022), p. 047. arXiv: [2205.01697 \[hep-ph\]](https://arxiv.org/abs/2205.01697).
- [57] Samuel Klein, Matthew Leigh, Stephen Mulligan, and Tobias Golling. “Strong CWoLa: Binary Classification Without Background Simulation.” In: *arXiv preprint* (2025). arXiv: [2503.14876 \[hep-ph\]](https://arxiv.org/abs/2503.14876).
- [58] Juan Antonio Aguilar-Saavedra and Francisco R. Joaquim. “Mass Unspecific Supervised Tagging (MUST) for Boosted Jets.” In: *JHEP* 2021.3 (2021), p. 012. arXiv: [2008.12792 \[hep-ph\]](https://arxiv.org/abs/2008.12792).
- [59] Shreecheta Chowdhury, Amit Chakraborty, and Saunak Dutta. *Probes of Anomalous Events at LHC with Self-Organizing Maps*. 2025. arXiv: [2503.09247 \[hep-ph\]](https://arxiv.org/abs/2503.09247). URL: <https://arxiv.org/abs/2503.09247>.
- [60] A. Author and B. Author. “Unsupervised Event Classification with Graphs on Classical and Photonic Quantum Computers.” In: *arXiv preprint* (2021). arXiv: [2104.02092 \[hep-ph\]](https://arxiv.org/abs/2104.02092).
- [61] A. Hammad, Mihoko M. Nojiri, and Masahito Yamazaki. *Quantum similarity learning for anomaly detection*. 2024. arXiv: [2411.09927 \[hep-ph\]](https://arxiv.org/abs/2411.09927). URL: <https://arxiv.org/abs/2411.09927>.
- [62] Sulaiman Alvi, Christian W. Bauer, and Benjamin Nachman. “Quantum Anomaly Detection for Collider Physics.” In: *JHEP* 2023.2 (2023), p. 220. arXiv: [2206.08391 \[hep-ph\]](https://arxiv.org/abs/2206.08391).
- [63] Shuai Zhang, Ke-Xin Chen, and Ji-Chong Yang. “Detect Anomalous Quartic Gauge Couplings at Muon Colliders with Quantum Kernel K-Means.” In: *Eur. Phys. J. C* 85.4 (2025), pp. 1–10. arXiv: [2409.07010 \[hep-ph\]](https://arxiv.org/abs/2409.07010).

- [64] O. Atkinson, A. Bhardwaj, C. Englert, V. Ngairangbam, and M. Spannowsky. “Anomaly detection with convolutional graph neural networks.” In: *JHEP* 08 (2021), p. 080. doi: [10.1007/JHEP08\(2021\)080](https://doi.org/10.1007/JHEP08(2021)080). eprint: [2105.07988](https://arxiv.org/abs/2105.07988).
- [65] Benjamin Nachman and David Shih. “Anomaly Detection with Density Estimation.” In: *Phys. Rev. D* 101 (2020), p. 075042. doi: [10.1103/PhysRevD.101.075042](https://doi.org/10.1103/PhysRevD.101.075042).
- [66] A. Andreassen, B. Nachman, and D. Shih. “Simulation Assisted Likelihood-Free Anomaly Detection.” In: *Phys. Rev. D* 101 (2020), p. 095004. doi: [10.1103/PhysRevD.101.095004](https://doi.org/10.1103/PhysRevD.101.095004). eprint: [2001.05001](https://arxiv.org/abs/2001.05001).
- [67] Kyle Metzger, Lana Xu, Mia Sodini, Thea K. Arrestad, Katya Govorkova, Gaia Grosso, and Philip Harris. *Anomaly preserving contrastive neural embeddings for end-to-end model-independent searches at the LHC*. 2025. arXiv: [2502.15926 \[hep-ex\]](https://arxiv.org/abs/2502.15926). URL: <https://arxiv.org/abs/2502.15926>.
- [68] Gregor Kasieczka et al. “The LHC Olympics 2020: A Community Challenge for Anomaly Detection in High Energy Physics.” In: *SciPost Phys.* 10 (2021), p. 5. doi: [10.21468/SciPostPhys.10.1.005](https://doi.org/10.21468/SciPostPhys.10.1.005).
- [69] Thea Arrestad et al. “The Dark Machines Anomaly Score Challenge: Benchmark Data and Model Independent Event Classification for the Large Hadron Collider.” In: *SciPost Physics* 12.1 (Jan. 2022). ISSN: 2542-4653. doi: [10.21468/scipostphys.12.1.043](https://doi.org/10.21468/scipostphys.12.1.043). URL: <http://dx.doi.org/10.21468/SciPostPhys.12.1.043>.
- [70] CMS Collaboration. *W, Z, and Higgs bosons as portals to exotic physics*. Accessed: 2025-05-05. 2022. URL: <https://cms.cern/news/w-z-and-higgs-bosons-portals-exotic-physics>.
- [71] CMS Collaboration. *Particle-flow reconstruction and global event description with the CMS detector*. Tech. rep. CMS-PAS-PFT-09-001. CMS Physics Analysis Summary. CERN, 2009. URL: <https://cds.cern.ch/record/2270046>.
- [72] G. et al. (ATLAS Collaboration) Aad. “The ATLAS Experiment at the CERN Large Hadron Collider.” In: *JINST* 3 (2008), S08003. doi: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [73] “The CMS experiment at the CERN LHC.” In: *JINST* 3 (2008), S08004. doi: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [74] Hervé Bourlard and Yves Kamp. “Auto-association by multilayer perceptrons and singular value decomposition.” In: *Biological Cybernetics* 59.4-5 (1988), pp. 291–294. doi: [10.1007/BF00332918](https://doi.org/10.1007/BF00332918).
- [75] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. “Learning representations by back-propagating errors.” In: *Nature* 323.6088 (1986), pp. 533–536. doi: [10.1038/323533a0](https://doi.org/10.1038/323533a0).
- [76] Akshansh Bhardwaj, Souvik Ghosh, Kaustuv Roy, and Kuver Sinha. “Unsupervised Classification of Jets Using Graph Neural Networks.” In: *Phys. Rev. D* 104.11 (2021), p. 115009. doi: [10.1103/PhysRevD.104.115009](https://doi.org/10.1103/PhysRevD.104.115009).
- [77] Kunihiko Fukushima. “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position.” In: *Biological Cybernetics* 36.4 (1980), pp. 193–202. doi: [10.1007/BF00344251](https://doi.org/10.1007/BF00344251).

- [78] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. "Gradient-based learning applied to document recognition." In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324. DOI: [10.1109/5.726791](https://doi.org/10.1109/5.726791).
- [79] Diederik P. Kingma and Max Welling. "Auto-Encoding Variational Bayes." In: *arXiv preprint arXiv:1312.6114* (2013). Accepted at ICLR 2014. URL: <https://arxiv.org/abs/1312.6114>.
- [80] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. "Stochastic backpropagation and approximate inference in deep generative models." In: *Proceedings of the 31st International Conference on Machine Learning*. PMLR. 2014, pp. 1278–1286. URL: <http://proceedings.mlr.press/v32/rezende14.html>.
- [81] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is All You Need." In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017. URL: https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fb0d053c1c4a845aa-Paper.pdf.
- [82] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. "Dynamic graph CNN for learning on point clouds." In: *ACM Transactions on Graphics (TOG)* 38.5 (2019), pp. 1–12. DOI: [10.1145/3326362](https://doi.org/10.1145/3326362). URL: <https://arxiv.org/abs/1801.07829>.
- [83] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. "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 (July 2014). ISSN: 1029-8479. DOI: [10.1007/jhep07\(2014\)079](https://doi.org/10.1007/jhep07(2014)079). URL: [http://dx.doi.org/10.1007/JHEP07\(2014\)079](http://dx.doi.org/10.1007/JHEP07(2014)079).
- [84] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. "An introduction to PYTHIA 8.2." In: *Computer Physics Communications* 191 (June 2015), 159–177. ISSN: 0010-4655. DOI: [10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024). URL: <http://dx.doi.org/10.1016/j.cpc.2015.01.024>.
- [85] S. Agostinelli et al. "GEANT4: A simulation toolkit." In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: [10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [86] Rene Brun and Fons Rademakers. "ROOT — An object oriented data analysis framework." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1 (1997). New Computing Techniques in Physics Research V, pp. 81–86. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(97\)00048-X](https://doi.org/10.1016/S0168-9002(97)00048-X). URL: <https://www.sciencedirect.com/science/article/pii/S016890029700048X>.
- [87] Matthias Fey and Jan Eric Lenssen. "Fast Graph Representation Learning with PyTorch Geometric." In: *CoRR* abs/1903.02428 (2019). arXiv: [1903.02428](https://arxiv.org/abs/1903.02428). URL: [http://arxiv.org/abs/1903.02428](https://arxiv.org/abs/1903.02428).