

School of Physics and Astronomy  
Queen Mary University of London

# **Charged Higgs Boson Search Using Low-Level Particle Analysis with Machine Learning Techniques**

Volodymyr Drobot (220546393)

April 25, 2025

Supervisor: Prof. Ulla Blumenschein

SPA6776 Extended Independent Project  
30 Credit Units

Submitted in partial fulfilment of the requirements for the degree of  
BSc Physics from Queen Mary University of London

## **Declaration**

I hereby certify that this project report, which is approximately fourteen thousand words in length, has been written by me at the School of Physics and Astronomy, Queen Mary University of London, that all material in this dissertation which is not my own work has been properly acknowledged, and that it has not been submitted in any previous application for a degree.

Volodymyr Drobot (220546393)

## Abstract

This dissertation presents a novel approach to the search for charged Higgs bosons at the ATLAS experiment using low-level particle analysis with machine learning techniques. We focus on the  $H^+ \rightarrow W^+ h \rightarrow W^+ b\bar{b}$  decay channel, comparing the performance of Convolutional Neural Networks (CNNs) and Transformer-based models for discriminating signal from background and distinguishing between leptonic ( $\ell\nu b\bar{b}$ ) and hadronic ( $q\bar{q} b\bar{b}$ ) decay modes of the  $W$  boson. Our approach leverages raw particle four-vectors and types, demonstrating that deep learning models can effectively learn complex patterns in high-energy physics data without explicit feature engineering. Results show distinctive signal discrimination, with both architectures achieving AUC values above 0.96. The Transformer model performs optimally for the leptonic channel (AUC of 0.973, 92.32% signal efficiency), while CNNs excel in the hadronic channel (96.12% signal efficiency). Overall, the hadronic decay mode demonstrates consistently higher signal efficiency across both architectures compared to the leptonic mode, suggesting a promising strategy for enhancing the sensitivity of charged Higgs searches at the LHC.

# Acknowledgments

I would like to express my gratitude to Professor Blumenschein for her guidance and support throughout this project. Special thanks to Sid for providing the dataset, valuable insights into the analysis techniques used in the ATLAS experiment and guidance on training machine learning models. I am also grateful to the ATLAS Collaboration for making this research possible.

# Contents

<b>Acknowledgments</b>	<b>2</b>
<b>1 Introduction</b>	<b>8</b>
1.1 Context and Significance of the Search for Charged Higgs Bosons . . . . .	8
1.2 The Higgs Mechanism and Electroweak Symmetry Breaking . . . . .	10
1.3 Overview of the ATLAS Experiment . . . . .	12
1.4 Signal and Background Processes . . . . .	14
1.4.1 Signal Processes . . . . .	14
1.4.2 Experimental Signatures and Jet Physics . . . . .	15
1.4.3 Background Processes . . . . .	17
1.5 Research Objectives and Dissertation Structure . . . . .	18
<b>2 Dataset and Preprocessing</b>	<b>19</b>
2.1 Dataset Overview . . . . .	19
2.1.1 Signal and Background Composition . . . . .	20
2.1.2 Data Structure . . . . .	20
2.1.3 Low-Level Particle Features . . . . .	20
2.2 Initial Data Visualization . . . . .	21
2.2.1 Distributions of Raw Features . . . . .	21
2.2.2 Mass and MET Distributions . . . . .	23
2.3 Event Selection and Categorization . . . . .	25
2.3.1 Selection Strategy . . . . .	25
2.3.2 Truth Information . . . . .	25
2.4 Data Preprocessing for Neural Networks . . . . .	26
2.4.1 MET Selection Criteria . . . . .	26
2.4.2 Channel Distribution After Selection . . . . .	27
2.4.3 Event Weights and Class Balancing . . . . .	28
2.4.4 Model Related Processing . . . . .	31
2.4.5 Dataset Partitioning . . . . .	31
<b>3 Machine Learning Methodology</b>	<b>32</b>
3.1 Motivation and Context for Machine Learning in High-Energy Physics .	32
3.2 Neural Network Fundamentals . . . . .	33
3.2.1 Activation Functions . . . . .	34
3.2.2 Loss Functions for Classification . . . . .	35
3.2.3 Backpropagation Algorithm . . . . .	35
3.2.4 Optimization . . . . .	37
3.2.5 Regularization . . . . .	38
3.3 Deep Learning Architectures for Charged Higgs Search . . . . .	39

3.3.1	Problem Formulation . . . . .	39
3.3.2	Data Preparation for Neural Networks . . . . .	39
3.3.3	Convolutional Neural Network Architecture . . . . .	41
3.3.4	Transformer Architecture . . . . .	43
3.4	Training Procedure . . . . .	45
3.5	Performance Evaluation Metrics . . . . .	47
3.5.1	Standard Classification Metrics . . . . .	47
3.5.2	ROC Curve and AUC . . . . .	47
3.5.3	Physics-Motivated Metrics . . . . .	48
<b>4</b>	<b>Results and Performance Evaluation</b>	<b>49</b>
4.1	Model Training Assessment . . . . .	49
4.2	Overall Classification Performance . . . . .	50
4.3	Detailed Performance Analysis . . . . .	51
4.3.1	Confusion Matrices . . . . .	51
4.3.2	Prediction Score Distributions . . . . .	52
4.4	Comparative Analysis of Architectures and Channels . . . . .	53
4.5	Physics Interpretation and Summary . . . . .	54
<b>5</b>	<b>Conclusion and Future Work</b>	<b>55</b>
5.1	Summary of Findings . . . . .	55
5.2	Limitations and Challenges . . . . .	56
5.3	Future Directions . . . . .	56
5.4	Broader Impact . . . . .	58
5.5	Final Remarks . . . . .	58

# List of Figures

1.1	The Standard Model of particle physics, showing the fundamental particles including the six quarks (purple), six leptons (green), four gauge bosons (red), and the Higgs boson (yellow) [3]. . . . .	9
1.2	The Higgs potential, resembling a "Mexican hat." The minimum energy occurs not at the origin but in a ring of non-zero values. When the field settles at one specific point on this ring, it spontaneously breaks the electroweak symmetry [11]. . . . .	10
1.3	Cross-sectional view of the ATLAS detector showing its primary components. The detector records trajectories and energy deposits of particles produced in proton-proton collisions at the interaction point [15]. . . .	12
1.4	Feynman diagrams showing the production and decay of a charged Higgs boson through the $H^+ \rightarrow W^+ h \rightarrow W^+ b\bar{b}$ channel: (a) with leptonic W decay ( $W^+ \rightarrow \ell^+ \nu$ ), and (b) with hadronic W decay ( $W^+ \rightarrow q\bar{q}$ ). . . .	14
1.5	Illustration of how increasing Lorentz boost affects the collimation of decay products. Left: At low boost, decay products are well-separated and reconstructed as individual jets. Middle: With moderate boost, decay products become more collimated. Right: At high boost (relevant for heavy charged Higgs bosons), decay products are highly collimated and may be reconstructed as a single jet with substructure [21]. . . .	16
1.6	Feynman diagrams of major background processes: (a) s-channel $t\bar{t} \rightarrow l^+\nu_l b\bar{b}$ process (semileptonic decay), and (b) W+jets production. These Standard Model processes can produce similar final state signatures to the charged Higgs signal. . . . .	17
2.1	Raw distribution of the momentum components (px, py, pz) for signal (blue) and background (red) events. The signal distributions show broader ranges with larger standard deviations, consistent with the higher energy scale of charged Higgs decays. . . . .	22
2.2	Raw energy distribution of particles in signal (blue) and background (red) events. Signal events contain particles with higher energies on average (mean: 416.53 GeV vs. 216.27 GeV), consistent with the decay of a heavy resonance. . . . .	22
2.3	Distribution of particle types in signal and background events. Small-radius jets (type 4) dominate in both signal and background events, with signal events showing a higher proportion of large-radius jets (type 3), which often contain the merged decay products of boosted objects. . . .	23
2.4	Distribution of signal events by charged Higgs mass point. . . . .	24

2.5	Overall distribution of missing transverse energy (MET) across all events, shown on a logarithmic scale. The distribution peaks at lower values and exhibits a long tail extending to 1 TeV, typical of collider physics data where high-MET events are rare but physically significant. . . . .	24
2.6	Confusion matrix showing the relationship between selection category and truth decay mode for all events. The matrix reveals significant misclassification between decay modes and substantial signal loss, highlighting the need for more sophisticated classification techniques. . . . .	26
2.7	Distribution of missing transverse energy (MET) for background events (blue histogram) and signal events in the qqbb channel (orange line) and lvbb channel (green line). The vertical dashed red line indicates the applied MET cut at 30 GeV. Note the logarithmic scale on the y-axis, highlighting the long tail of high-MET events, especially in the signal lvbb channel. . . . .	27
2.8	Distribution of events between the leptonic (lvbb) and hadronic (qqbb) decay channels after selection criteria are applied. The leptonic channel contains significantly more events (887,701 signal, 753,618 background) than the hadronic channel (300,245 signal, 230,697 background), despite the W boson's higher hadronic branching ratio in nature. . . . .	28
3.1	Schematic representation of an artificial neuron. Multiple inputs ( $X_1, X_2, \dots, X_n$ ) are multiplied by corresponding weights ( $W_1, W_2, \dots, W_n$ ), summed with a bias term ( $b$ ), and passed through an activation function (in this case $\sigma$ ) to produce the output [33]. . . . .	33
3.2	Architecture of a fully connected neural network with one hidden layer. Each connection represents a weight parameter that is learned during training. . . . .	34
3.3	Illustration of gradient descent optimization. Starting from a random initial value, each learning step moves the parameters ( $w$ ) in the direction of steepest descent until reaching a minimum of the cost function [37]. . . . .	37
3.4	Illustration of dropout regularization: (a) Standard Neural Network with all neurons active, and (b) Network after applying dropout with randomly deactivated neurons (marked with X) [41]. . . . .	38
3.5	Architecture of the Convolutional Neural Network model with attention mechanism used for charged Higgs identification. The model processes particle sequences through convolutional layers followed by an attention mechanism that identifies the most relevant particles. . . . .	41
3.6	Architecture of the Transformer model used for charged Higgs identification. Each particle is treated as a token in a sequence, enabling the self-attention mechanism to capture global relationships between particles regardless of their positions. . . . .	43
4.1	Learning curves for the lvbb channel models showing training loss (blue) and validation loss (orange) as functions of training epoch. . . . .	49
4.2	Learning curves for the qqbb channel models showing training loss (blue) and validation loss (orange) as functions of training epoch. . . . .	50

4.3	ROC curves for CNN (blue) and Transformer (red) models on both decay channels. The AUC values are 0.967 (CNN) and 0.973 (Transformer) for lvbb, and 0.966 (CNN) and 0.964 (Transformer) for qqbb. . . . .	50
4.4	Confusion matrices for the lvbb channel models. Values represent the percentage of events in each category after applying a classification threshold of 0.5. . . . .	51
4.5	Confusion matrices for the qqbb channel models. Values represent the percentage of events in each category after applying a classification threshold of 0.5. . . . .	52
4.6	Prediction score distributions for signal (blue) and background (red) events in the lvbb channel, comparing CNN (left) and Transformer (right) models. The vertical dashed line indicates the default classification threshold of 0.5. . . . .	52
4.7	Prediction score distributions for signal (blue) and background (red) events in the qqbb channel, comparing CNN (left) and Transformer (right) models. The vertical dashed line indicates the default classification threshold of 0.5. . . . .	53

# Chapter 1

## Introduction

### 1.1 Context and Significance of the Search for Charged Higgs Bosons

The year 2012 marked a milestone in particle physics with the discovery of a neutral Higgs boson at the Large Hadron Collider (LHC) [1, 2]. This breakthrough confirmed the existence of the fundamental field that gives elementary particles their mass, completing the Standard Model (SM) after decades of theoretical prediction and experimental search. The Higgs mechanism, proposed in the 1960s, explains how particles acquire mass through interactions with a quantum field that permeates all space. The observed particle, with a mass of approximately 125 GeV, represents a quantum excitation of this Higgs field and was the final missing piece needed to validate the SM’s theoretical framework.

The Standard Model of particle physics is our most comprehensive theory of the fundamental constituents of matter and their interactions. Developed in the latter half of the 20th century, it classifies all known elementary particles into two main categories: fermions (matter particles with half-integer spin) and bosons (force-carrying particles with integer spin). The fermions include six quarks (up, down, charm, strange, top, bottom) and six leptons (electron, muon, tau, and their corresponding neutrinos), while the bosons include the photon (mediating electromagnetism), W and Z bosons (mediating weak nuclear force), and gluons (mediating strong nuclear force) [3]. The Standard Model explains how these particles interact through three of the four fundamental forces—electromagnetic, weak, and strong—with remarkable precision, successfully predicting experimental results across many orders of magnitude in energy [4].

Despite its remarkable successes, the SM is not considered a complete theory and has several limitations. It does not incorporate gravity, explain the existence of dark matter or dark energy, account for neutrino masses, or address the hierarchy problem—the question of why the Higgs boson mass is so much lighter than the Planck scale [5]. These limitations strongly motivate the exploration of theories beyond the Standard Model (BSM).

Among the most compelling extensions to the SM are those that predict an expanded Higgs sector, which often includes charged Higgs bosons ( $H^\pm$ ). Many well-motivated theoretical frameworks, such as Two-Higgs-Doublet Models (2HDMs) [6] and Minimal Supersymmetric Standard Model (MSSM) [7], naturally predict the existence of charged Higgs bosons as part of a richer scalar sector. The search for charged

## Standard Model of Elementary Particles

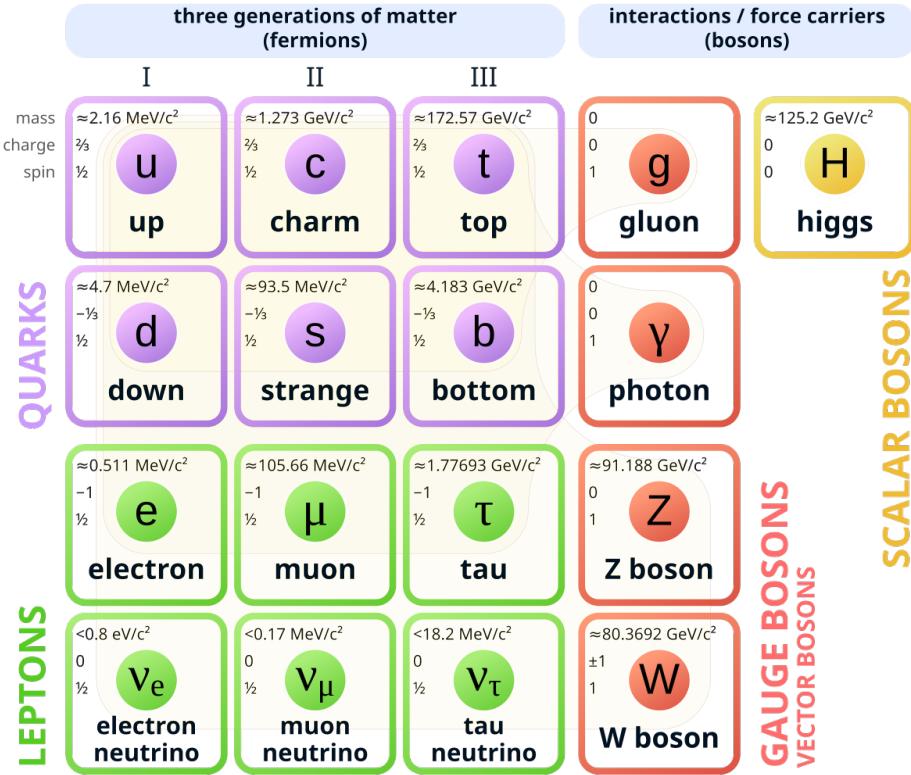


Figure 1.1: The Standard Model of particle physics, showing the fundamental particles including the six quarks (purple), six leptons (green), four gauge bosons (red), and the Higgs boson (yellow) [3].

Higgs bosons represents one of the most direct experimental approaches to probe BSM physics and could potentially address some of the outstanding questions of particle physics.

For instance, charged Higgs bosons could provide additional sources of CP violation beyond what is available in the SM. CP violation, the asymmetry between matter and antimatter under the combined operation of charge conjugation (C) and parity inversion (P), is a crucial ingredient in explaining the observed matter-antimatter asymmetry in the universe, a phenomenon for which the SM provides an insufficient explanation. Discovery of charged Higgs bosons could also play a significant role in dark matter physics or contribute to the mechanisms that generate neutrino masses [7].

From a broader theoretical perspective, charged Higgs bosons are not exclusive to 2HDMs and the MSSM. They appear in various other compelling BSM frameworks, including models with additional symmetries, composite Higgs models, and models with extra spatial dimensions. The fundamental properties of charged Higgs bosons—their masses, their interactions with other particles (couplings), and their preferred ways of decaying (branching ratios)—can offer valuable clues about the underlying BSM theory and help physicists distinguish between different theoretical possibilities [8].

Therefore, the search for charged Higgs bosons is of paramount significance in high-energy physics. Their discovery would open a new window into the fundamental structure of matter and energy, potentially resolving some of the most pressing questions

that the Standard Model leaves unanswered and guiding us towards a more complete understanding of the universe.

## 1.2 The Higgs Mechanism and Electroweak Symmetry Breaking

The Standard Model unifies electromagnetic and weak interactions through the electroweak theory, based on an  $SU(2)_L \times U(1)_Y$  gauge symmetry [9]. This symmetry, while mathematically elegant, requires all force-carrying particles (gauge bosons) to be massless—contradicting experimental observations of massive  $W$  and  $Z$  bosons. The Higgs mechanism resolves this discrepancy by introducing a complex scalar field that undergoes spontaneous symmetry breaking [10].

This scalar field is structured as a complex  $SU(2)$  doublet containing four degrees of freedom. These four degrees arise because each of the two components of the doublet is complex, with both real and imaginary parts that can vary independently. Before symmetry breaking, these four independent components correspond to four distinct massless scalar fields that can each be excited separately. The key insight of the Higgs mechanism is that the potential energy of this field has a shape resembling a "Mexican hat"—with its minimum not at zero but at a ring of non-zero values, as illustrated in Figure 1.2.

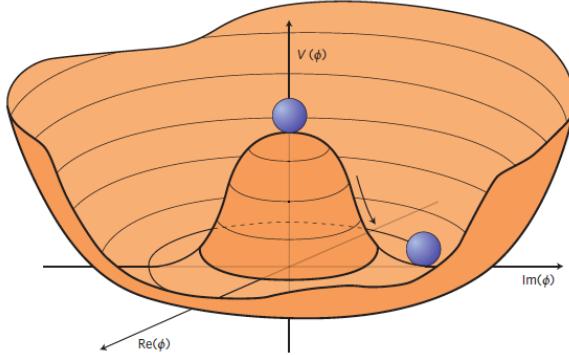


Figure 1.2: The Higgs potential, resembling a "Mexican hat." The minimum energy occurs not at the origin but in a ring of non-zero values. When the field settles at one specific point on this ring, it spontaneously breaks the electroweak symmetry [11].

When the temperature of the universe decreased, the field settles into one specific point on this minimum-energy ring, spontaneously breaking the original symmetry. This process establishes a vacuum expectation value (VEV) of approximately 246 GeV throughout space [12].

Once electroweak symmetry breaks in this manner, three of the four degrees of freedom in the Higgs field are absorbed by the  $W^+$ ,  $W^-$ , and  $Z$  bosons, giving them mass and providing their longitudinal polarization components. The remaining degree of freedom manifests as the physical Higgs boson, experimentally confirmed in 2012. The photon remains massless because the unbroken  $U(1)_{EM}$  subgroup does not couple to the Higgs field.

This implementation with a single  $SU(2)$  doublet represents the simplest possible mechanism for electroweak symmetry breaking, but there is no fundamental principle

that requires the Higgs sector to be minimal. The choice appears somewhat arbitrary from a theoretical perspective, and extended scalar sectors represent some of the most straightforward and theoretically well-motivated extensions to the Standard Model.

The simplest extension of the SM Higgs sector is the Two-Higgs-Doublet Model (2HDM), which introduces a second  $SU(2)$  doublet of complex scalar fields. After electroweak symmetry breaking, this structure results in five physical Higgs bosons: two neutral scalars ( $h$  and  $H$ , where  $h$  is typically identified with the observed 125 GeV boson), one neutral pseudoscalar ( $A$ , a particle with spin 0 that changes sign under a parity transformation), and two charged Higgs bosons ( $H^\pm$ ). The 2HDM framework encompasses several different types, characterized by how the Higgs doublets couple to fermions [6]:

- **Type I:** All fermions couple to the same Higgs field, with the other doublet having no direct fermion couplings.
- **Type II:** Up-type quarks couple to one Higgs field, while down-type quarks and charged leptons couple to the other. This structure appears naturally in the Minimal Supersymmetric Standard Model (MSSM).
- **Type III (Lepton-specific):** Quarks couple to one doublet while leptons couple to the other.
- **Type IV (Flipped):** Up-type quarks and leptons couple to one doublet, while down-type quarks couple to the other.

In supersymmetric extensions of the SM, particularly the MSSM, an extended Higgs sector is not just a possibility but a requirement for theoretical consistency. Supersymmetry, which proposes a symmetry between fermions and bosons, requires at least two Higgs doublets to give masses to both up-type and down-type fermions while maintaining anomaly cancellation. The properties of the charged Higgs bosons in the MSSM depend primarily on two parameters: the ratio of vacuum expectation values of the two doublets ( $\tan \beta$ ) and the mass of the pseudoscalar Higgs boson ( $m_A$ ), which in many scenarios is closely related to the charged Higgs mass [13].

While these theoretical frameworks provide compelling motivation for the existence of charged Higgs bosons, their discovery requires experimental verification. Experimental evidence necessitates sophisticated detection systems capable of identifying unique signatures of these particles in high-energy collisions. This is where experiments at high-energy particle colliders, such as the Large Hadron Collider, become essential for the exploration of physics beyond the Standard Model.

## 1.3 Overview of the ATLAS Experiment

The ATLAS (A Toroidal LHC ApparatuS) detector is one of the two general-purpose detectors at the LHC, designed to investigate a wide range of physics phenomena including searches for new particles such as charged Higgs bosons. Spanning 46 meters in length and 25 meters in diameter, ATLAS consists of multiple specialized subdetector systems arranged in concentric layers around the beam pipe [14]. When protons collide at the center of the detector, the energy transformation creates particles that propagate outward through these successive detection layers, allowing ATLAS to record their trajectories and energy deposits.

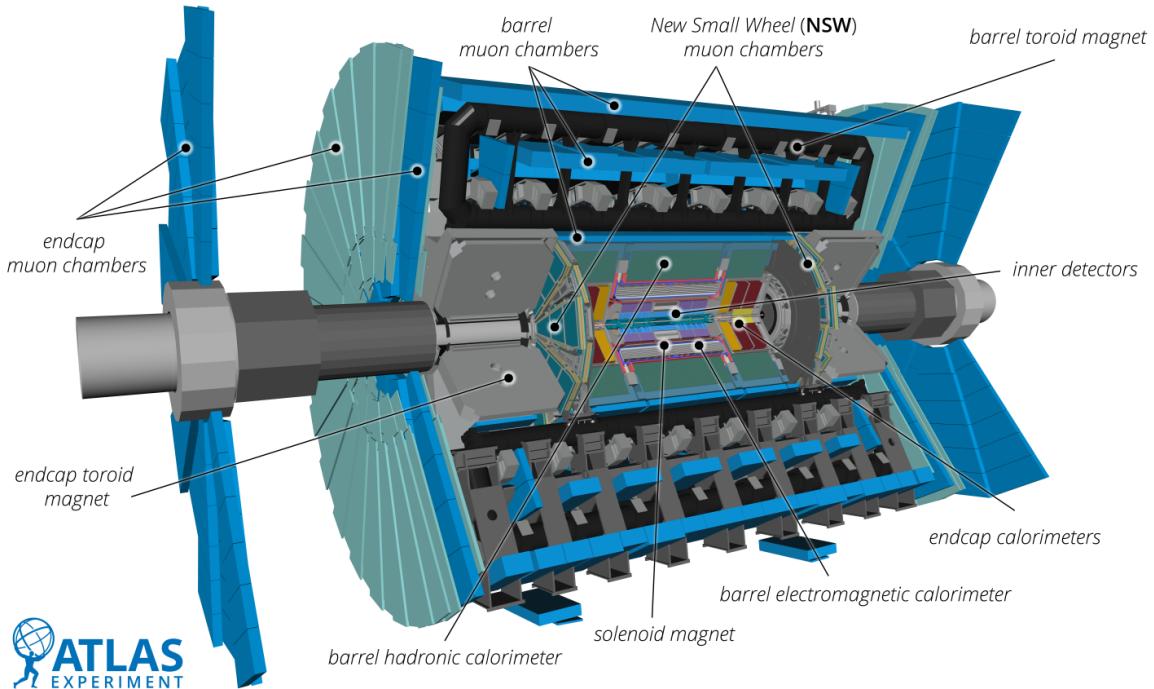


Figure 1.3: Cross-sectional view of the ATLAS detector showing its primary components. The detector records trajectories and energy deposits of particles produced in proton-proton collisions at the interaction point [15].

The ATLAS trigger and data acquisition system is designed to select approximately 1,000 events per second for permanent storage from the 40 million proton-proton collisions occurring each second during LHC operation. This selection occurs through a sophisticated multi-level trigger system that rapidly identifies events of potential physics interest, ensuring that rare processes like charged Higgs production can be captured despite their low production cross-sections.

The Inner Detector tracks charged particles as they traverse a magnetic field. This tracking system comprises three technologies: silicon pixel detectors positioned nearest to the collision point for high-resolution position measurements, silicon strip detectors for additional tracking coordinates, and a transition radiation tracker which aids electron identification through radiation emission at material interfaces.

The calorimeter system measures particle energies through complete absorption. The electromagnetic calorimeter utilizes liquid argon technology to measure electrons and photons via electromagnetic shower development. The hadronic calorimeter, constructed with steel and scintillating tiles, measures jets from quarks and gluons as they

fragment into particle showers through a process called hadronization (the formation of hadrons from quarks and gluons).

The muon spectrometer constitutes the outermost detection layer. Muons, which penetrate the inner detector systems with minimal energy loss, leave a distinct signal in the muon spectrometer, providing crucial information for physics analyses.

The magnet system enables momentum measurements through particle path curvature. The system consists of a central solenoid generating a 2 Tesla field surrounding the Inner Detector, and toroidal magnets around the muon system producing a variable field for muon trajectory analysis [14].

This detection configuration allows ATLAS to reconstruct and identify the various particle types produced in proton-proton collisions:

- **Electrons:** fundamental particles with a negative charge and relatively small mass.
- **Muons:** heavier cousins of electrons, also negatively charged, and capable of penetrating more detector material.
- **Jets:** sprays of particles produced by the hadronization of quarks and gluons, the fundamental constituents of protons.
- **Missing transverse energy (MET):** inferred from an imbalance in momentum, signals the presence of neutrinos, which interact weakly and escape detection, or potentially new, undetected particles.

For the charged Higgs analysis in this dissertation, the detector's ability to precisely reconstruct jets and identify b-jets (b-tagging) [16] is particularly important for identifying the b-quarks from the  $h \rightarrow b\bar{b}$  decay. Similarly, lepton identification and missing energy resolution are crucial for reconstructing the W boson decay products in the leptonic channel.

The low-level particle information used in this analysis directly benefits from the high-resolution tracking and calorimetry provided by the ATLAS detector. For each event, we analyze the raw particle four-vectors ( $p_x$ ,  $p_y$ ,  $p_z$ ,  $E$ ) and particle type information, capturing the fundamental characteristics of the collisions.

This dissertation analyzes simulated data from the ATLAS experiment, focusing on the decay channel  $H^+ \rightarrow W^+ h \rightarrow W^+ b\bar{b}$ , where the W boson decays either leptonically (lvbb) or hadronically (qqbb). The dataset includes simulated signal events for charged Higgs mass points from 800 GeV to 3000 GeV, and relevant Standard Model background processes.

Understanding the detailed particle-level signatures of both signal and background events is essential for developing effective discrimination techniques. The following section describes these processes in detail.

## 1.4 Signal and Background Processes

Searching for charged Higgs bosons at the LHC requires a thorough understanding of both the signal signature and the Standard Model processes that can mimic it. This section describes the specific signal topology investigated in this dissertation and the background processes that pose challenges for its identification. By characterizing these processes at the particle level, we establish the foundation for the low-level particle analysis approach that forms the core of this work.

### 1.4.1 Signal Processes

The analysis presented in this dissertation focuses specifically on the charged Higgs boson production and decay via the channel  $H^+ \rightarrow W^+ h \rightarrow W^+ b\bar{b}$ , where the  $W$  boson decays through two distinct modes, leading to different experimental signatures.

$$H^+ \rightarrow W^+ h \rightarrow W^+ b\bar{b} \quad (1.1)$$

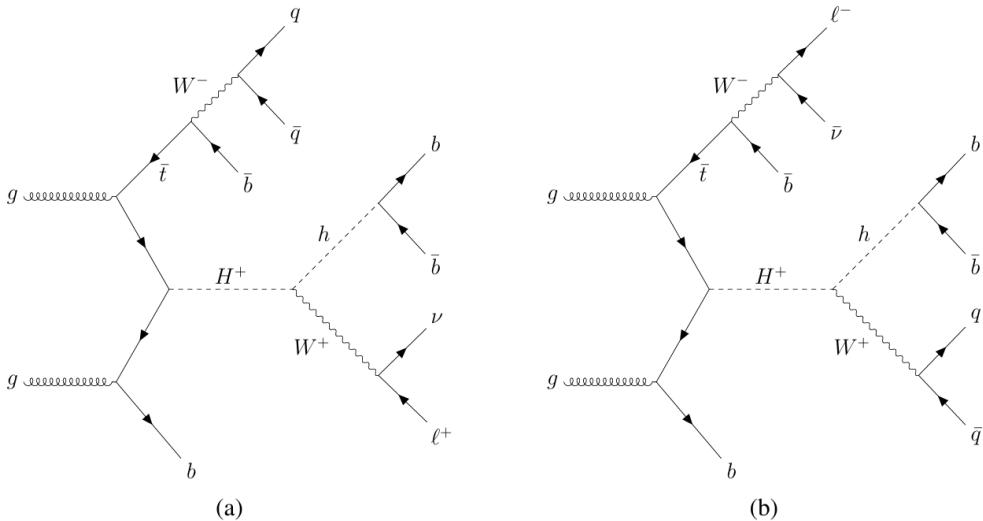


Figure 1.4: Feynman diagrams showing the production and decay of a charged Higgs boson through the  $H^+ \rightarrow W^+ h \rightarrow W^+ b\bar{b}$  channel: (a) with leptonic  $W$  decay ( $W^+ \rightarrow \ell^+ \nu$ ), and (b) with hadronic  $W$  decay ( $W^+ \rightarrow q\bar{q}$ ).

- **Leptonic decay (lvbb):**  $H^+ \rightarrow W^+ h \rightarrow \ell^+ \nu \ell^+ \bar{b} b$ . The  $W$  boson decays to a lepton (electron or muon) and a neutrino. This produces a final state containing one lepton, missing transverse energy (from the neutrino), and  $b$ -jets from the SM Higgs decay.
- **Hadronic decay (qqbb):**  $H^+ \rightarrow W^+ h \rightarrow q\bar{q}' b\bar{b}$ . The  $W$  boson decays to a pair of quarks. This results in an all-hadronic final state with  $b$ -jets from the Higgs decay and additional jets from the  $W$  boson decay.

### 1.4.2 Experimental Signatures and Jet Physics

Both decay channels critically depend on our ability to accurately identify and characterize jets formed by quarks from the decay chain. In particular, the identification of b-jets from the Higgs decay to b-quark pairs is essential for distinguishing signal from background, while the proper reconstruction of jets from the W boson’s hadronic decay is vital for the qqbb channel. For the lvbb channel, accurate lepton identification and missing energy reconstruction are equally important to capture the W boson’s leptonic decay products. To understand how these physical processes manifest in the detector, it’s important to examine how quarks materialize as jets in high-energy collisions.

In high-energy particle collisions, quarks and gluons cannot exist freely due to color confinement, a fundamental principle of quantum chromodynamics (QCD). Instead, they undergo a process known as hadronization, forming collimated sprays of particles called jets [17]. Jets are complex objects consisting of hadrons, which are composite particles made up of quarks held together by the strong force. In the context of detector physics, jets appear as clusters of energy deposits in calorimeters accompanied by tracks in the inner detector.

The nature of jets is particularly important for this analysis. When a quark or gluon is produced with substantial energy, it initiates a parton shower through successive radiation of gluons and quark-antiquark pair production [18]. This cascade of particles subsequently hadronizes, resulting in a collimated cone of particles traveling in approximately the same direction as the original parton. The properties of these jets—their energy, momentum, mass, and substructure—provide crucial information about the originating particles and processes.

A key distinction in this analysis is between b-jets and light-flavor jets. B-jets originate from the b-quarks in the  $h \rightarrow b\bar{b}$  decay and have distinctive characteristics due to the relatively long lifetime of B-hadrons (approximately  $1.5 \times 10^{-12}$  seconds) [19]. This allows them to travel a measurable distance before decaying, creating secondary vertices and decay products with specific momentum and multiplicity patterns. Identifying these b-jets is critical for our signal process, as the Higgs to b-quark decay is a defining characteristic that must be distinguished from backgrounds, particularly those involving top quarks.

For the heavy charged Higgs masses considered in this study (800-3000 GeV), the decay products receive significant Lorentz boosts due to the large mass difference between the initial and final states. This boost has profound implications for the detector signatures [20]:

- **Collimated decay products:** As particles become more boosted, their decay products become increasingly collimated in the laboratory frame due to the Lorentz transformation. For example, the two b-quarks from the Higgs decay may become so close that they are reconstructed as a single large-radius jet rather than two separate jets.
- **Enhanced jet substructure:** These boosted jets contain rich substructure that reflects the underlying decay topology. The angular distributions, energy sharing, and multiplicity patterns within the jet provide valuable information about whether it originated from a heavy resonance decay or from QCD background processes.

- **High transverse momentum ( $\mathbf{p}_T$ ):** The decay products carry large momenta, particularly in the transverse plane perpendicular to the beam axis. High- $\mathbf{p}_T$  jets and leptons serve as important triggers and selection criteria for identifying potential signal events.

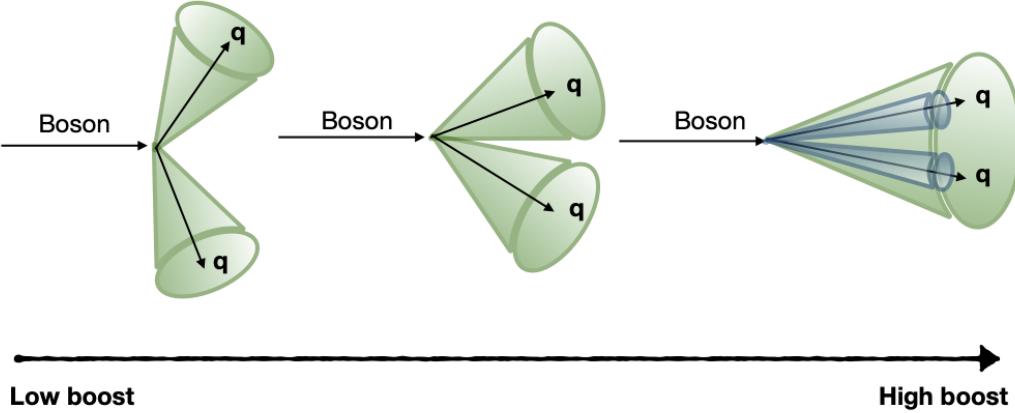


Figure 1.5: Illustration of how increasing Lorentz boost affects the collimation of decay products. Left: At low boost, decay products are well-separated and reconstructed as individual jets. Middle: With moderate boost, decay products become more collimated. Right: At high boost (relevant for heavy charged Higgs bosons), decay products are highly collimated and may be reconstructed as a single jet with substructure [21].

In experimental analyses, jets are typically reconstructed using clustering algorithms such as anti- $k_T$  with specific radius parameters [22]. For capturing the entire decay products of boosted objects, large-radius jets (often with  $R = 0.8$  or larger) are employed, while for resolving individual b-jets in less boosted scenarios, smaller radius jets (typically  $R = 0.4$ ) are used. This analysis leverages both approaches as appropriate for different kinematic regimes.

At these high mass points, several distinctive characteristics emerge that can be exploited for signal identification:

- Highly energetic and boosted decay products with distinctive momentum and angular distributions
- Collimated b-jet pairs from the Higgs decay, potentially forming a single jet with characteristic two-prong substructure
- Characteristic angular correlations between the Higgs decay products and the W boson decay products
- High-momentum leptons (in the lvbb channel) or high-energy jets (in the qqbb channel) from the W boson decay

Effective identification of the signal events requires a thorough understanding of the Standard Model background processes, which can exhibit complex and overlapping features. The following section describes these background processes in detail, outlining the challenges they present for accurate signal discrimination.

### 1.4.3 Background Processes

Despite the distinctive signature of charged Higgs production, several Standard Model processes can mimic this signal and constitute significant backgrounds for this analysis. Effectively discriminating signal from these backgrounds is the central challenge addressed in this dissertation.

The main background processes for this analysis are:

- **Top quark pair production ( $t\bar{t}$ ):** This constitutes the dominant background, as it produces final states with W bosons and b-jets that can closely mimic the signal topology. When both top quarks decay via  $t \rightarrow Wb$  with the subsequent decay of one W boson to leptons and the other to quarks, the final state particles closely resemble those from the signal process.
- **Single top production:** Similar to  $t\bar{t}$  but with a single top quark, this process can produce W bosons and b-jets that contribute significantly to the background.
- **W+jets production:** W bosons produced in association with jets can mimic the signal when some jets are misidentified as b-jets. This is particularly relevant for the leptonic decay channel.
- **Z+jets production:** Z bosons produced with jets can contribute to the background, particularly when the Z decays to neutrinos (providing missing energy) or heavy-flavor quarks.
- **Diboson production (WW, WZ, ZZ):** These processes can produce multiple bosons and jets that may resemble the signal signature.

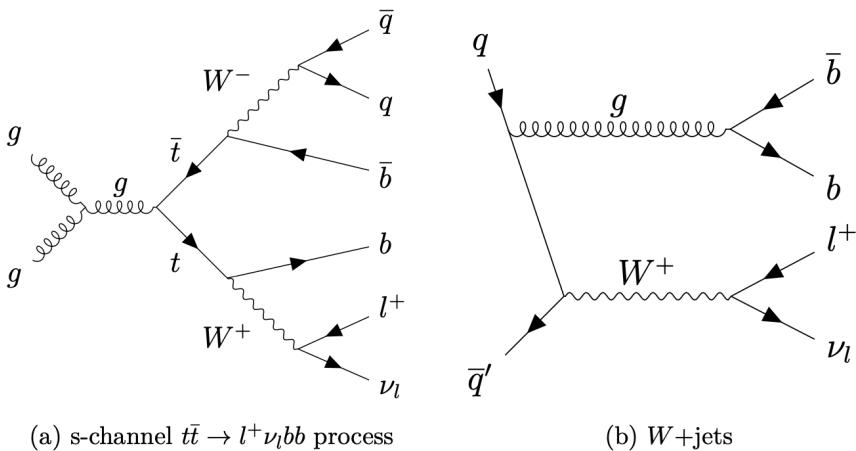


Figure 1.6: Feynman diagrams of major background processes: (a) s-channel  $t\bar{t} \rightarrow l^+\nu_l b\bar{b}$  process (semileptonic decay), and (b)  $W+$ jets production. These Standard Model processes can produce similar final state signatures to the charged Higgs signal.

Standard Model processes like  $t\bar{t}$  production involve W bosons and b-jets that closely resemble the  $H^+ \rightarrow W^+ h \rightarrow W^+ b\bar{b}$  signal, making traditional cut-based approaches insufficient. However, several subtle kinematic differences can be exploited:

- Differences in the distributions of jet  $p_T$  and energy
- Invariant mass distributions of particle combinations
- Angular separations between particles
- Correlations among multiple particles that may not be obvious from high-level variables

These subtle kinematic differences, while difficult to leverage with traditional methods, can potentially be exploited through advanced analysis techniques. Traditional cut-based analyses relying on high-level reconstructed variables may not fully exploit these differences at the particle level, limiting their ability to effectively separate signal and background, especially in complex final states with multiple jets. Furthermore, factors such as detector resolution, background modeling uncertainties, and the decreasing signal cross-section at higher masses contribute to the difficulty of these searches.

## 1.5 Research Objectives and Dissertation Structure

To address these challenges and potentially enhance search sensitivity, this dissertation explores the application of machine learning techniques to low-level particle data. The primary objectives are to develop and evaluate models for identifying charged Higgs boson events, comparing the performance of different neural network architectures—particularly Convolutional Neural Networks (CNNs) and Transformers. This work assesses the discrimination power between signal and background events, as well as between different decay modes ( $l\nu bb$  vs.  $qqbb$ ), determining which channel offers the best prospects for charged Higgs discovery. The research also interprets the results in the context of LHC searches, estimating potential sensitivity improvement and identifying promising regions of parameter space for future exploration.

The remainder of this dissertation is structured as follows:

- **Chapter 2: Dataset and Preprocessing** describes the ATLAS experiment and the dataset used in this analysis, including the signal and background samples, event selection criteria, and preprocessing steps.
- **Chapter 3: Machine Learning Methodology** presents the machine learning approach developed for this analysis, including the problem formulation, neural network architectures, training procedure, and evaluation metrics.
- **Chapter 4: Results and Performance Evaluation** analyzes the performance of the models for charged Higgs identification, comparing architectures and decay channels to identify the most promising approaches.
- **Chapter 5: Conclusion and Future Work** summarizes the findings, discusses the limitations of the current approach, and outlines directions for future research in this field.

The dissertation aims to demonstrate the potential of low-level particle analysis with machine learning techniques to enhance the sensitivity of charged Higgs searches, potentially opening new avenues for the discovery of physics beyond the Standard Model.

# Chapter 2

## Dataset and Preprocessing

This chapter describes the dataset used in this analysis, the event selection criteria, and the preprocessing steps applied to prepare the data for machine learning. The dataset consists of simulated events from the ATLAS experiment, representing both signal (charged Higgs boson) and background processes. Rather than using high-level reconstructed variables, this analysis focuses on low-level particle information—specifically the four-momenta and types of particles in each event—to investigate whether deep learning models can identify patterns that might be missed in traditional approaches. We begin by providing an overview of the dataset’s composition and structure, including a visualization of the unprocessed data. Then we describe the event selection and preprocessing steps applied to prepare the data for machine learning in subsequent parts of this chapter.

The complete code implementation for this analysis, including all data preprocessing, model architectures, training procedures and evaluation metrics, is available in a public [GitHub repository](#). The repository contains comprehensive documentation of the codebase and instructions for reproducing the results presented in this dissertation.

### 2.1 Dataset Overview

The dataset used in this study consists of simulated proton-proton collision events modeling charged Higgs boson production and decay, along with various Standard Model background processes. These events were generated through a comprehensive simulation pipeline that accurately models both the underlying physics processes and the ATLAS detector response.

Monte Carlo simulations are essential in particle physics as they enable the study of processes that are too rare to observe frequently in real data and provide "truth" information about the actual identity of particles, which is crucial for training supervised machine learning models. For our analysis, signal and background events were first produced using specialized Monte Carlo event generators implementing theoretical models for charged Higgs production and Standard Model processes, creating four-vectors of all final-state particles from simulated proton-proton collisions.

To realistically represent how these particles would be measured in the ATLAS detector, the generated events were processed through detailed detector simulation. For charged Higgs signal samples with masses below 500 GeV, a parameterized fast simulation (ATLFASTII) was employed to efficiently generate the necessary statistics while maintaining acceptable accuracy. For the higher mass points of 500 GeV and

above—which are the focus of this analysis—a full GEANT4-based simulation of the ATLAS detector was used [23]. This comprehensive approach models particle interactions with detector material at a microscopic level, accounting for energy deposition, multiple scattering, and secondary particle production.

The simulated detector responses were then processed through the same reconstruction algorithms used for actual ATLAS data. This approach ensures our machine learning models are trained on data that faithfully represents the experimental challenges present in a real charged Higgs search, including detector resolution effects, reconstruction inefficiencies, and background contamination patterns specific to the ATLAS experiment.

### 2.1.1 Signal and Background Composition

The dataset contains two types of events: signal events representing the production and decay of charged Higgs bosons through the channel  $H^+ \rightarrow W^+ h \rightarrow W^+ b\bar{b}$ , with the  $W$  boson decaying either leptonically ( $\ell^+ \nu_\ell$ ) or hadronically ( $q\bar{q}'$ ), and background events representing Standard Model processes that can mimic the signal signature. Signal events are assigned Dataset IDs (DSIDs) in the range 510115-510124, with each DSID corresponding to a specific charged Higgs mass point, while background events have DSIDs below 500,000 or above 600,000, including processes such as  $t\bar{t}$  production, single top production, W+jets, Z+jets, and diboson production.

### 2.1.2 Data Structure

The dataset is stored in ROOT format, a specialized file format for high-energy physics data, organized in a tree structure. For each event, we extract the low-level particle features (px, py, pz, E, and particle type for each particle), event-level variables (DSID, truth\_W\_decay\_mode, selection\_category, eventWeight), and Missing transverse energy (MET). These features are read using the custom `read_file` function from the `read_low_level.py` module, which handles the conversion from the ROOT format to numpy arrays suitable for machine learning applications.

The structure of our input data reflects the hierarchical nature of particle physics events:

- Each event contains a variable number of particles
- Each particle is characterized by its four-momentum components and type
- Additional event-level variables provide context and selection criteria

### 2.1.3 Low-Level Particle Features

The core of this analysis is the use of low-level particle information as input to machine learning models. For each particle, we extract five key features:

- **px**: The x-component of the particle’s momentum
- **py**: The y-component of the particle’s momentum
- **pz**: The z-component of the particle’s momentum

- **E**: The energy of the particle
- **type**: An integer encoding the particle type

These low-level features provide a complete kinematic description of each particle, preserving the fundamental physical information without relying on derived quantities that might obscure important correlations. By using these raw features, we allow the neural network to discover the most relevant combinations and transformations directly from the data.

The particle type feature is encoded as an integer where:

- -1: Padding (not a real particle)
- 0: Electron
- 1: Muon
- 2: Neutrino
- 3: Large-radius jet
- 4: Small-radius jet

This encoding scheme allows the model to distinguish between different types of particles, which is crucial since different particles interact differently with the detector and play different roles in the decay cascade of the charged Higgs boson.

## 2.2 Initial Data Visualization

To understand the dataset’s initial properties, this section presents visualizations of the unprocessed data, before any selection or preprocessing steps are applied. These visualizations reveal the inherent structure of the data and highlight differences between signal and background events, as well as between leptonic and hadronic decay channels.

### 2.2.1 Distributions of Raw Features

Figure 2.1 shows the distribution of the momentum components ( $p_x$ ,  $p_y$ ,  $p_z$ ) for both signal and background events. These distributions reveal several distinctive characteristics. Signal events on average demonstrate significantly wider spreads with standard deviations approximately 1.5-3 times larger than background events (302.46 GeV, 301.36 GeV, and 412.80 GeV for  $p_x$ ,  $p_y$ , and  $p_z$  in signal events compared to 120.27 GeV, 120.68 GeV, and 245.99 GeV for background). Both  $p_x$  and  $p_y$  distributions are approximately symmetric around zero (with mean values close to zero: -0.22 GeV and -0.29 GeV for signal; -0.16 GeV and 0.04 GeV for background), reflecting the cylindrical symmetry of the detector around the beam axis. The  $p_z$  distribution shows a broader spread in both directions, for both the signal and background distributions, since momentum along the beam axis is not constrained by detector acceptance in the same way as transverse components. The difference in  $z$  momentum indicates higher longitudinal momentum transfers consistent with the decay of a heavy resonance [24].

These distribution characteristics fundamentally reflect the physical differences between signal and background processes. The signal processes involve the decay of

massive charged Higgs bosons (800-3000 GeV), which transfer substantial momentum to their decay products. In contrast, background processes typically involve Standard Model interactions at lower energy scales, resulting in less energetic particles with narrower momentum distributions.

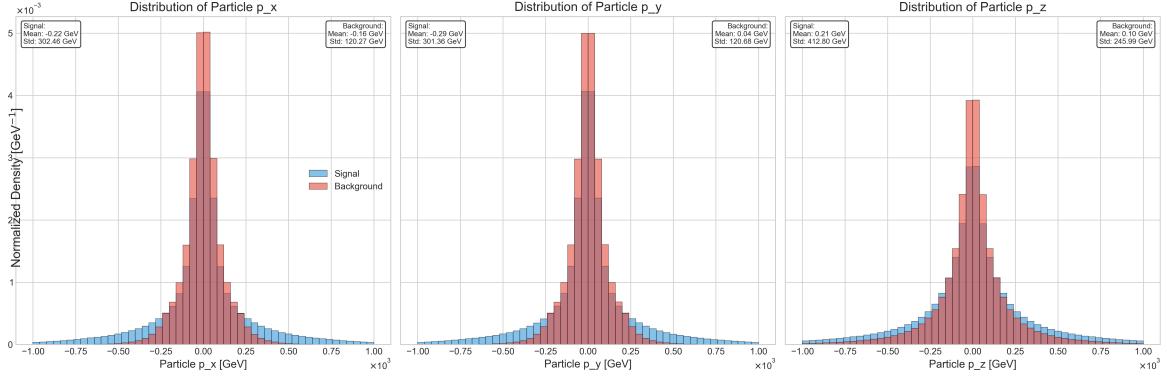


Figure 2.1: Raw distribution of the momentum components ( $p_x$ ,  $p_y$ ,  $p_z$ ) for signal (blue) and background (red) events. The signal distributions show broader ranges with larger standard deviations, consistent with the higher energy scale of charged Higgs decays.

Similarly, Figure 2.2 presents the raw energy distribution of particles in signal and background events. The signal events clearly contain particles with higher energies on average, with a mean energy of 416.53 GeV compared to 216.27 GeV for background events. The signal distribution also exhibits a more pronounced high-energy tail extending beyond 1.5 TeV, while the background distribution falls off more rapidly after about 500 GeV. The standard deviation of the energy distribution is also much larger for signal events (428.53 GeV) compared to background events (211.15 GeV), indicating a wider spread of energy values.

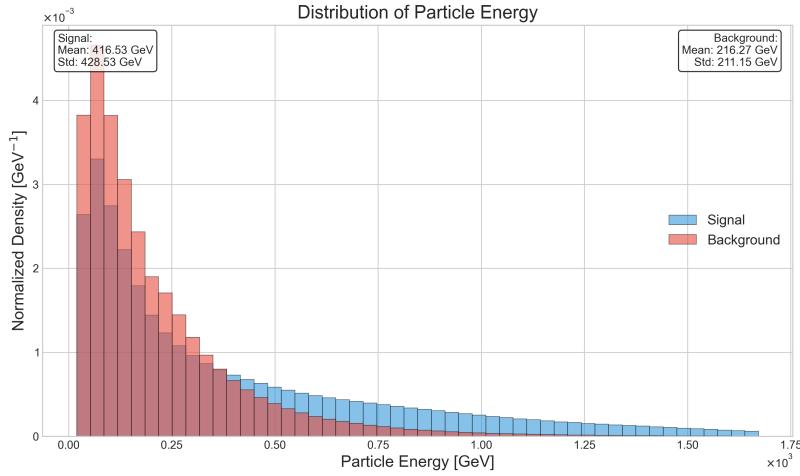


Figure 2.2: Raw energy distribution of particles in signal (blue) and background (red) events. Signal events contain particles with higher energies on average (mean: 416.53 GeV vs. 216.27 GeV), consistent with the decay of a heavy resonance.

This clear separation in energy distributions represents one of the most discriminating features between signal and background, highlighting the potential for effective

classification using these low-level features. The energy tail behavior is particularly important for identifying charged Higgs events, as the decay of such massive particles naturally produces high-energy decay products that populate the tail region where background is suppressed [8].

Figure 2.3 reveals that small-radius jets dominate the particle composition, constituting approximately 67-68% of particles in both signal (19,723,832) and background (38,609,964) events, as shown by the exact counts in the figure. Neutrinos form the second most common particle type (11-12%), followed by large-radius jets (9-10%), and leptons (electrons and muons) each comprising 5-6% of particles. The relative proportions are remarkably similar between signal and background events, though background contains proportionally more neutrinos and large-radius jets. This distribution reflects the typical particle content of high-energy proton-proton collisions, where QCD processes producing jets dominate the final states [25].

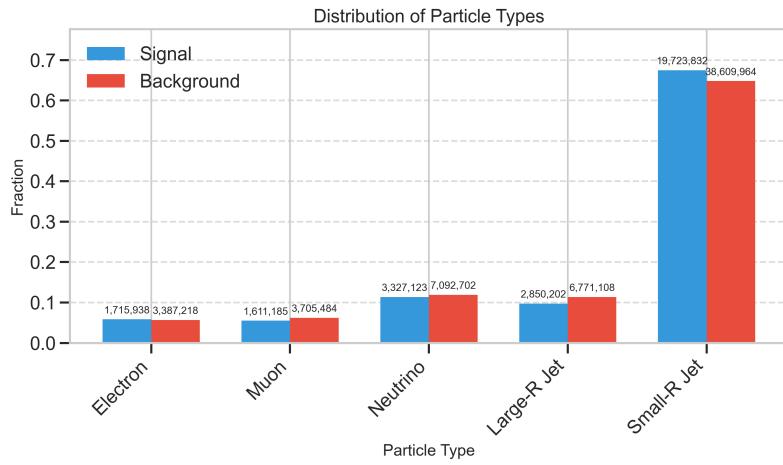


Figure 2.3: Distribution of particle types in signal and background events. Small-radius jets (type 4) dominate in both signal and background events, with signal events showing a higher proportion of large-radius jets (type 3), which often contain the merged decay products of boosted objects.

## 2.2.2 Mass and MET Distributions

Figure 2.4 shows the distribution of signal events across different charged Higgs mass points from 800 GeV to 3000 GeV. Interestingly, the event counts across mass points are remarkably uniform, ranging from approximately 297,190 events at 800 GeV to a maximum of 345,540 events at 1600 GeV. This uniform distribution does not reflect the natural production cross-section, which would decrease significantly with increasing mass [26]. Instead, it indicates a deliberate simulation strategy to ensure adequate statistics across all mass points, with higher masses being oversampled relative to their expected production rates. This approach is standard in high-energy physics simulations to ensure sufficient events for training machine learning models across the entire mass range of interest.

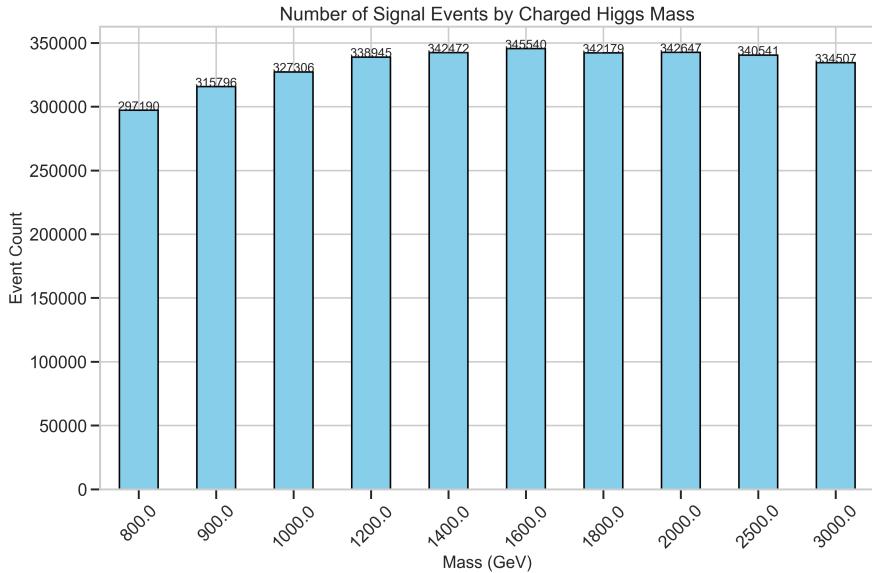


Figure 2.4: Distribution of signal events by charged Higgs mass point.

Figure 2.5 presents the overall distribution of missing transverse energy (MET) across all events in the dataset. The distribution follows a characteristic exponential decay pattern typical of MET in high-energy collisions, with a peak around 50-100 GeV and a long tail extending beyond 1 TeV. This missing energy arises primarily from neutrinos, which escape the detector without interacting, and is a crucial variable for event selection in the lvbb channel where a neutrino from the W boson decay contributes significant MET [27]. The logarithmic scale on the y-axis highlights the rarity of very high MET events, which often correspond to highly boosted decay products from massive resonances.

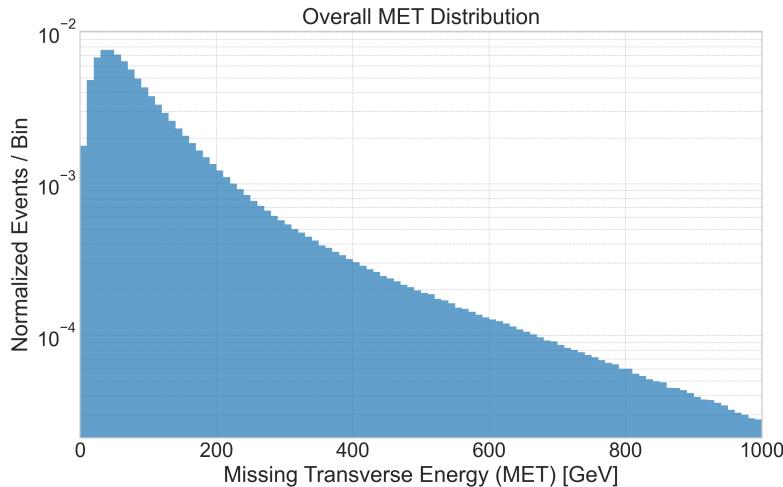


Figure 2.5: Overall distribution of missing transverse energy (MET) across all events, shown on a logarithmic scale. The distribution peaks at lower values and exhibits a long tail extending to 1 TeV, typical of collider physics data where high-MET events are rare but physically significant.

These visualizations provide a baseline understanding of the raw data characteristics before any selection or preprocessing steps are applied. In the next section, we will

examine in detail how signal and background events are identified and selected, which is a crucial first step before applying more sophisticated analysis methods.

## 2.3 Event Selection and Categorization

In high-energy physics analyses, event selection plays a crucial role in isolating the signal from the background. This section describes the selection criteria applied to our dataset and the categorization of events into different decay channels, which is a critical first step before applying more sophisticated machine learning techniques.

### 2.3.1 Selection Strategy

The event selection is performed based on the `selection_category` variable, which encodes the result of a preliminary classification of events based on their detector signature. This pre-selection is done by the ATLAS software and uses a combination of criteria to identify potential signal candidates. Events with `selection_category` values of 0, 8, or 10 are classified as potential lvbb candidates, typically having an identified lepton, substantial missing transverse energy (from the neutrino), and b-jets from the Higgs decay. Events with `selection_category` values of 3 or 9 are classified as potential qqbb candidates, typically having multiple jets, including b-jets from the Higgs decay.

### 2.3.2 Truth Information

While the `selection_category` variable is used to classify events based on their detector signature, the `truth_W_decay_mode` variable provides the actual decay mode of the W boson for signal events. This truth information comes from the simulation and is not available in real detector data, but is essential for training supervised machine learning models. For signal events, `truth_W_decay_mode = 1` indicates that the W boson decayed leptonically (lvbb), while `truth_W_decay_mode = 2` indicates that it decayed hadronically (qqbb). For background events, the `truth_W_decay_mode` value is not meaningful, as there is no charged Higgs or W boson from its decay.

Figure 2.6 shows the relationship between the selection category and the truth decay mode for all events, including signal and background, in the form of a confusion matrix. This matrix reveals several important aspects of our dataset:

- The lvbb selection correctly identifies 887,701 true lvbb events, but also incorrectly selects 201,742 true qqbb events and 1,138,067 background ("other") events.
- The qqbb selection correctly identifies 300,245 true qqbb events, but also incorrectly selects 98,759 true lvbb events and 185,862 background events.
- A substantial number of true signal events (676,799 lvbb and 488,740 qqbb) are categorized as "other," representing potential signal loss in the selection.
- The largest category by far is background events correctly identified as "other" (6,441,910 events), reflecting the efficient rejection of clear background cases.

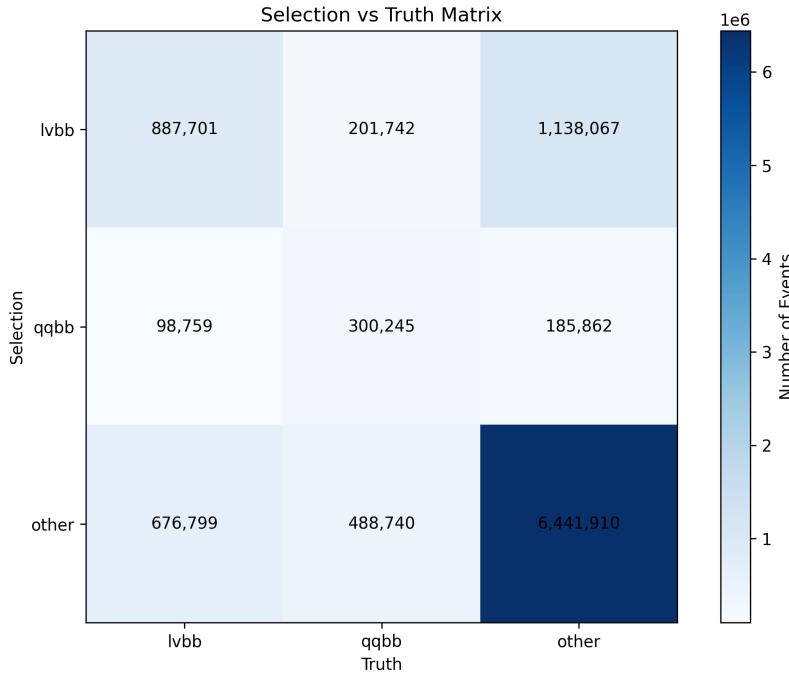


Figure 2.6: Confusion matrix showing the relationship between selection category and truth decay mode for all events. The matrix reveals significant misclassification between decay modes and substantial signal loss, highlighting the need for more sophisticated classification techniques.

This confusion matrix illustrates the challenges in event selection and the potential for improvement through machine learning techniques. The substantial misclassification rates between decay modes and the significant signal loss highlight the need for more sophisticated approaches to event selection and classification.

## 2.4 Data Preprocessing for Neural Networks

Machine learning models, particularly neural networks, require careful preprocessing of the input data to achieve optimal performance. The preprocessing steps must preserve the physical information in the data while transforming it into a format suitable for neural network training. This section describes these preprocessing steps in detail.

### 2.4.1 MET Selection Criteria

Figure 2.7 compares the Missing Transverse Energy (MET) distribution, previously introduced in Section 2.2, across the different event categories: background events (blue histogram) versus signal events in both the qqbb channel (orange line) and lvbb channel (green line). This comparison highlights key differences used to optimize event selection. Several important features are evident in this distribution:

- The background events (blue) are heavily concentrated at low MET values, with a sharp peak below 50 GeV.

- Signal events in the lvbb channel (green) show a much broader distribution extending to higher MET values, which is expected due to the neutrino from the  $W \rightarrow \ell\nu$  decay.
- Signal events in the qqbb channel (orange) show an intermediate distribution, with more high-MET events than background but fewer than the lvbb channel.
- All distributions show long tails extending to 1000 GeV and beyond, reflecting the high energy scales involved in these processes.

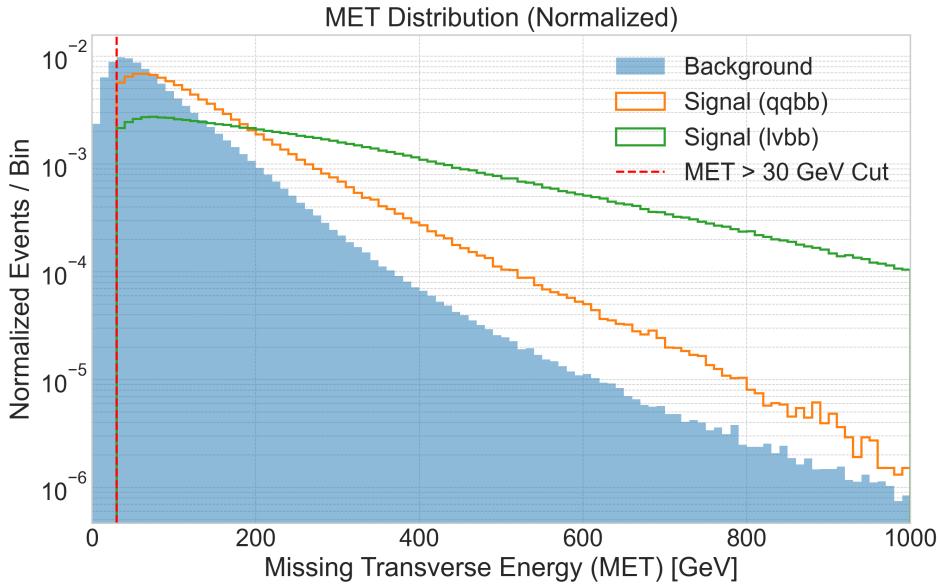


Figure 2.7: Distribution of missing transverse energy (MET) for background events (blue histogram) and signal events in the qqbb channel (orange line) and lvbb channel (green line). The vertical dashed red line indicates the applied MET cut at 30 GeV. Note the logarithmic scale on the y-axis, highlighting the long tail of high-MET events, especially in the signal lvbb channel.

Based on these distributions, we apply a selection criterion requiring  $\text{MET} > 30 \text{ GeV}$  (indicated by the vertical red dashed line) to enhance signal purity. This cut provides significant discrimination power across both decay channels, removing a substantial portion of background events while preserving the majority of signal. The effectiveness stems from fundamental topological differences: signal events feature complex decay chains naturally producing larger MET values, while background processes predominantly occupy the low-MET region where apparent missing energy arises primarily from detector resolution effects [19]. This selection strategy ensures cleaner samples for our analysis while maintaining a consistent approach across both channels.

#### 2.4.2 Channel Distribution After Selection

Figure 2.8 shows the distribution of events between the leptonic (lvbb) and hadronic (qqbb) decay channels after applying the MET cut and other selection criteria. The analysis of this distribution reveals several interesting features about our dataset composition. The leptonic channel substantially dominates the event count, containing

887,701 signal and 753,618 background events, compared to just 300,245 signal and 230,697 background events in the hadronic channel. This distribution appears counterintuitive given that the hadronic decay mode of the W boson has a significantly higher branching ratio in nature (approximately 67% compared to 33% for the leptonic mode) [28].

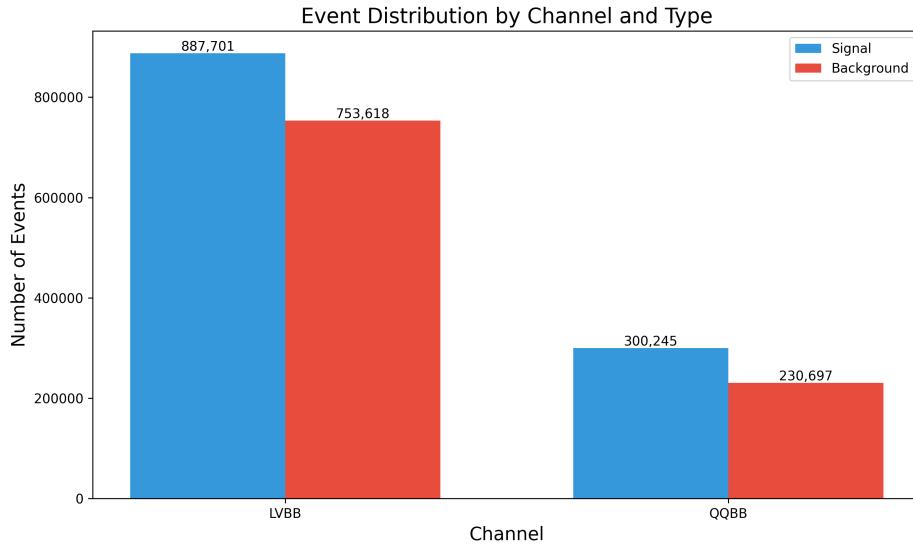


Figure 2.8: Distribution of events between the leptonic (lvbb) and hadronic (qqbb) decay channels after selection criteria are applied. The leptonic channel contains significantly more events (887,701 signal, 753,618 background) than the hadronic channel (300,245 signal, 230,697 background), despite the W boson’s higher hadronic branching ratio in nature.

Despite having fewer events, the qqbb channel exhibits a slightly better signal-to-background ratio of 1.30, compared to 1.18 in the lvbb channel. The disproportionately higher event count in the lvbb channel likely reflects a deliberate strategy in the simulation process to prioritize this channel due to its cleaner experimental signature in real detector conditions, despite its lower natural branching fraction. Leptons provide clear signatures that are easier to identify against the complex hadronic background of proton-proton collisions, making them valuable triggers for potential charged Higgs events. This distribution informs our approach to train separate models for each decay channel, allowing us to optimize for their distinct features while maintaining a sufficient number of events in each for effective training. The channel-specific approach enables the models to learn the unique characteristics of each decay mode without forcing them to distinguish between fundamentally different event topologies.

### 2.4.3 Event Weights and Class Balancing

After categorizing events into their respective channels, handling event weights appropriately becomes crucial for training effective machine learning models. Event weights in high-energy physics account for various factors including cross-section differences between processes, detector acceptance effects, and sampling considerations in the Monte Carlo simulation. Without proper weighting, the model would learn a biased representation that doesn’t reflect the true physical proportions of signal and background events in nature.

For the lvbb channel, we implemented a two-step weighting approach. First, we preserve the original physics-motivated event weights from the simulation that account for cross-section differences and detector effects. Then, we apply an additional class balancing factor that equalizes the total weighted contributions of signal and background events. As demonstrated by the preprocessing statistics presented in Figure 2.8, the lvbb channel contained 887,701 signal events and 753,618 background events, giving a raw signal-to-background ratio of approximately 1.18. However, when considering the original event weights, the total signal weight was 428,097.51 while the background weight was only 16,138.32, revealing a significant imbalance in the weighted contributions.

To address this imbalance, we compute adjustment factors  $\alpha_s$  and  $\alpha_b$  for signal and background events, respectively:

$$W_s = \sum_{i \in \text{signal}} w_i^{\text{orig}} \quad (2.1)$$

$$W_b = \sum_{i \in \text{background}} w_i^{\text{orig}} \quad (2.2)$$

$$\alpha_s = 1.0 \quad (2.3)$$

$$\alpha_b = \frac{W_s}{W_b} \quad (2.4)$$

where  $W_s$  and  $W_b$  are the total original weights for signal and background events,  $w_i^{\text{orig}}$  is the original event weight for event  $i$ , and  $\alpha_s$  and  $\alpha_b$  are the adjustment factors. The adjusted weights  $w_i^{\text{adj}}$  are then calculated as:

$$w_i^{\text{adj}} = \begin{cases} w_i^{\text{orig}} \cdot \alpha_s & \text{if event } i \text{ is signal} \\ w_i^{\text{orig}} \cdot \alpha_b & \text{if event } i \text{ is background} \end{cases} \quad (2.5)$$

In our specific case,  $\alpha_s = 1.0$  and  $\alpha_b \approx 26.5$  (the ratio of total signal weight to total background weight). This resulted in perfectly balanced effective signal-to-background weight ratio of exactly 1.0, where both classes contribute equally to the model training despite their different event counts:

$$\sum_{i \in \text{signal}} w_i^{\text{adj}} = \sum_{i \in \text{background}} w_i^{\text{adj}} \quad (2.6)$$

Importantly, this approach preserves the relative weights within each class, maintaining the physics relationships between different signal processes and between different background processes, while only adjusting the overall scale between the two classes.

This balancing strategy is particularly important for the lvbb channel due to its significant background contamination (as seen in the earlier confusion matrix analysis, where only 39.9% of selected lvbb events are true signal). Without such balancing, the model training would be biased toward correctly identifying the more heavily weighted signal events at the expense of properly handling background rejection.

For the qqbb channel, the weighting approach is similarly thorough but adapted to its different class distribution. Here, we calculated class weights inversely proportional

to class frequencies in the dataset and applied these weights as multiplicative factors to the original physics-based event weights. The preprocessing statistics show that the qqbb channel contained 300,245 signal events and 230,697 background events.

For the qqbb channel, we used a standard inverse frequency approach to calculate the class weights:

$$n_s = \text{number of signal events} \quad (2.7)$$

$$n_b = \text{number of background events} \quad (2.8)$$

$$N = n_s + n_b \text{ (total number of events)} \quad (2.9)$$

$$c = \text{number of classes (2 in our case)} \quad (2.10)$$

$$\beta_s = \frac{N}{c \cdot n_s} \quad (2.11)$$

$$\beta_b = \frac{N}{c \cdot n_b} \quad (2.12)$$

where  $\beta_s$  and  $\beta_b$  are the class weight factors for signal and background, respectively. The adjusted weights are then calculated as:

$$w_i^{\text{adj}} = \begin{cases} w_i^{\text{orig}} \cdot \beta_s & \text{if event } i \text{ is signal} \\ w_i^{\text{orig}} \cdot \beta_b & \text{if event } i \text{ is background} \end{cases} \quad (2.13)$$

Based on our event distribution, we computed class weight factors of  $\beta_b = 1.15073451$  for background and  $\beta_s = 0.88418125$  for signal events. This inverse frequency approach typically causes less dramatic changes to the total event weights compared to the direct balancing method used for lvbb, while still effectively countering potential bias from class imbalance.

This approach is well-suited for the qqbb channel, which has a higher signal purity (51.4% as shown earlier) and less extreme differences in the original event weights between signal and background classes.

The different weighting strategies for the two channels are justified by their distinct characteristics. First, the lvbb channel has more background contamination, necessitating a more direct approach to equalizing the total signal and background weights. Second, the lvbb channel contains significantly more events (1,641,319 compared to 530,942 for qqbb), making it especially important to ensure balanced learning. Third, the background composition differs between channels, with the lvbb channel containing more diverse background sources (including ttbar, W+jets) while the qqbb channel faces primarily QCD multijet backgrounds. Finally, both approaches preserve the relative weights within each class, ensuring that the physical cross-section differences between processes within the signal or background categories are maintained.

The total sum of event weights before balancing in the lvbb channel was 444,235.83, which increased to 856,195.03 after the balancing procedure. This increase reflects the scaling applied to background events to match the larger total weight of signal events. For the qqbb channel, the class weighting approach adjusted the influence of each class proportionally to their inverse frequencies without dramatically changing the overall weight scale.

In both cases, the adjusted weights ensure that the neural networks learn to identify signal events effectively despite their relative rarity compared to background processes.

During model training, these weights are directly incorporated into the loss function calculation, which will be discussed in detail in Chapter 3 of dissertation.

The effectiveness of these weighting strategies will be reflected in the performance metrics of the trained models, particularly in terms of signal efficiency and background rejection across different mass points. By ensuring proper weighting, we prevent the inherent class imbalance of high-energy physics data from biasing the model's predictions, leading to more reliable and physically meaningful results.

#### 2.4.4 Model Related Processing

To prepare the data for machine learning, several key preprocessing steps were applied. These steps ensure the data is in a suitable format for neural network training while preserving the essential physical information:

- **Particle Count Standardization:** Since different events contain different numbers of particles, we standardized the input by setting a maximum number of particles per event and applying appropriate padding.
- **Categorical Encoding:** Particle types were encoded as integers, with special handling to distinguish between physical particles and padding.

The detailed implementation of these preprocessing techniques and their specific configurations will be discussed in Chapter 3, where we describe the machine learning methodology in depth.

#### 2.4.5 Dataset Partitioning

The preprocessed data was split into three distinct subsets to ensure robust model development and evaluation:

- **Training set (70%):** Used to train the models by adjusting their parameters through optimization
- **Validation set (15%):** Used for hyperparameter tuning and early stopping decisions during training
- **Test set (15%):** Used only for final performance evaluation, providing an unbiased assessment of the models

The splitting was performed with stratification to maintain the same signal-to-background ratio across all sets. Additionally, we ensured that different charged Higgs mass points were represented proportionally in each set, which is crucial for ensuring that the models generalize across the entire mass range of interest.

This comprehensive preprocessing and partitioning pipeline transforms the raw particle physics data into a format suitable for deep learning while preserving the essential physical information and accounting for the unique characteristics of high-energy physics data. Through careful event selection, feature scaling, and appropriate data splitting, we have prepared distinct datasets for the lvbb and qqbb channels, creating a solid foundation for the machine learning methodology described in the following chapter.

# Chapter 3

## Machine Learning Methodology

This chapter presents the machine learning approaches developed for charged Higgs boson identification using low-level particle information. We begin by discussing the motivation and context for applying machine learning in high-energy physics, followed by an examination of the fundamental concepts of neural networks. We then describe in detail the specific architectures implemented for this analysis, the training procedures employed, and the metrics used to evaluate performance.

### 3.1 Motivation and Context for Machine Learning in High-Energy Physics

High-energy physics (HEP) experiments like those at the Large Hadron Collider generate enormous volumes of complex data. The ATLAS detector alone produces approximately one petabyte of raw data per second during operation [14], which is filtered through a multi-level trigger system to select events of potential interest. Even after this initial filtering, the remaining data presents significant analytical challenges due to:

- The complex, high-dimensional nature of particle collision events
- Subtle kinematic differences between signal and background processes
- Low signal-to-background ratios, particularly for rare processes like charged Higgs production
- The need to accurately model intricate detector responses and experimental uncertainties

Traditional analysis approaches in HEP rely on manually engineered high-level features and cut-based selection criteria. While these methods have proven successful for many physics analyses, they may not fully exploit the rich information content of the raw detector data [29]. Furthermore, they require extensive domain knowledge and can inadvertently discard potentially valuable information during the feature engineering process.

Machine learning, particularly deep learning, offers an alternative approach that can potentially address these limitations. Deep neural networks can process high-dimensional data directly, automatically learning relevant features and complex pat-

terns without explicit programming. In recent years, machine learning has demonstrated increasing success in various HEP applications, including jet classification [30], event selection [31], and anomaly detection [32].

This dissertation explores the application of deep learning techniques to the search for charged Higgs bosons, focusing specifically on low-level particle information analysis. Rather than relying on derived high-level features, we train neural networks directly on the basic kinematic properties of particles (four-momenta and particle types), allowing the models to discover discriminative patterns that might be missed in traditional analyses. We compare two state-of-the-art neural network architectures (Convolutional Neural Networks (CNNs) and Transformer models), assess their strengths and potential for enhancing the sensitivity of charged Higgs searches.

### 3.2 Neural Network Fundamentals

Neural Networks (NNs) are computational models inspired by the structure and function of biological neural networks. They consist of interconnected processing units called neurons, organized into layers: an input layer that receives data, one or more hidden layers that perform computations, and an output layer that produces the final prediction. The network learns by adjusting the connection strengths (weights) between neurons to minimize the difference between outputs and ground truth.

The fundamental building block of a neural network is the artificial neuron, illustrated in Figure 3.1. An artificial neuron:

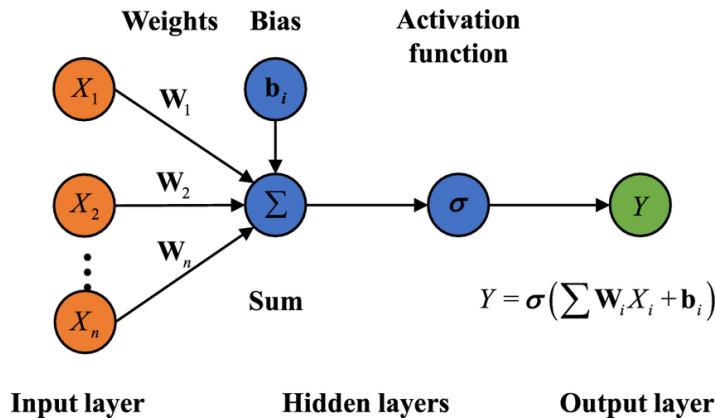


Figure 3.1: Schematic representation of an artificial neuron. Multiple inputs ( $X_1, X_2, \dots, X_n$ ) are multiplied by corresponding weights ( $W_1, W_2, \dots, W_n$ ), summed with a bias term ( $b$ ), and passed through an activation function (in this case  $\sigma$ ) to produce the output [33].

- Receives multiple input signals ( $X_1, X_2, \dots, X_n$ )
  - Assigns weights ( $W_1, W_2, \dots, W_n$ ) to these inputs
  - Computes a weighted sum plus a bias term:  $z = \sum_{i=1}^n W_i X_i + b_i$
  - Applies a non-linear activation function (in this case sigmoid):  $Y = \sigma(z)$
  - Outputs the result for subsequent processing

When multiple neurons are connected in layers, they form a network capable of learning complex functions, as shown in Figure 3.2. The input layer neurons receive the raw data features, hidden layer neurons perform intermediate computations, and output layer neurons produce the final prediction [34].

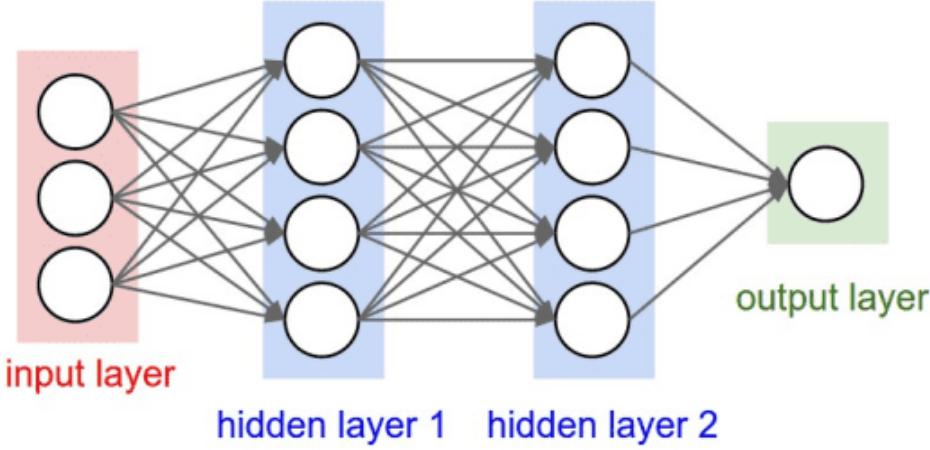


Figure 3.2: Architecture of a fully connected neural network with one hidden layer. Each connection represents a weight parameter that is learned during training.

Information propagates through the network in what is known as a forward pass. For each layer, the output of the previous layer serves as input, and the computation follows the pattern described for the individual neuron: weighted sum followed by activation function application. This process continues until the output layer produces the network's prediction.

### 3.2.1 Activation Functions

Activation functions are critical components of neural networks as they introduce non-linearity, enabling the network to learn complex, non-linear relationships in the data. Without activation functions, neural networks would be equivalent to linear models, regardless of their depth. Several common activation functions are used in modern neural networks:

The sigmoid function transforms inputs to values between 0 and 1, making it suitable for binary classification problems:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (3.1)$$

The hyperbolic tangent ( $\tanh$ ) function is similar to sigmoid but maps inputs to values between -1 and 1, providing stronger gradients:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3.2)$$

The Rectified Linear Unit (ReLU) function is one of the most widely used activation functions in modern neural networks due to its computational efficiency and effectiveness in mitigating the vanishing gradient problem:

$$\text{ReLU}(x) = \max(0, x) \quad (3.3)$$

The softmax function is typically used in the output layer for multi-class classification problems, as it normalizes outputs into a probability distribution:

$$\text{softmax}(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (3.4)$$

For our binary classification task of identifying charged Higgs events, we use the sigmoid activation function in the output layer, as it naturally produces values between 0 and 1 that can be interpreted as probabilities.

### 3.2.2 Loss Functions for Classification

Loss functions quantify the difference between predicted values and ground truth, providing the objective to be minimized during training. For binary classification tasks like ours, the Binary Cross-Entropy (BCE) loss is the standard choice. BCE measures the difference between the predicted probability distribution and the true distribution:

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (3.5)$$

where  $y_i$  is the true label (0 or 1),  $\hat{y}_i$  is the predicted probability, and  $N$  is the number of samples. BCE loss penalizes confident incorrect predictions more severely than less confident ones, encouraging the model to assign high probabilities to the correct class.

In high-energy physics, different events often have different importance based on their physics cross-sections, detector acceptance, and other factors, which are captured in event weights. To account for these weights, we use Weighted Binary Cross-Entropy:

$$L_{\text{weighted}} = -\frac{1}{N} \sum_{i=1}^N w_i [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (3.6)$$

where  $w_i$  is the weight for sample  $i$ . This ensures that the model learns to prioritize events according to their physics-based importance, rather than treating all events equally.

### 3.2.3 Backpropagation Algorithm

Backpropagation is the cornerstone algorithm that enables neural networks to learn from data. It provides an efficient way to compute gradients of the loss function with respect to all network parameters, which are then used to update these parameters through gradient descent.

The algorithm consists of two main phases: a forward pass and a backward pass. During the forward pass, input data propagates through the network layer by layer, generating predictions as described in the beginning of Section 3.2. During the backward pass, the error between predictions and true values is calculated and propagated backwards through the network to determine how each parameter contributed to this error.

Mathematically, for a simple network with one hidden layer, the backpropagation process works as follows:

1. **Forward Pass:** For an input  $\mathbf{x}$ , compute the hidden layer activations:

$$\mathbf{h} = f(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}) \quad (3.7)$$

where  $\mathbf{W}^{(1)}$  is the weight matrix connecting the input to the hidden layer,  $\mathbf{b}^{(1)}$  is the bias vector, and  $f$  is the activation function. Then compute the output:

$$\hat{\mathbf{y}} = g(\mathbf{W}^{(2)}\mathbf{h} + \mathbf{b}^{(2)}) \quad (3.8)$$

where  $\mathbf{W}^{(2)}$  and  $\mathbf{b}^{(2)}$  are the weights and biases of the output layer, and  $g$  is the output activation function.

2. **Error Calculation:** Compute the loss  $L(\hat{y}, y)$  between the prediction  $\hat{y}$  and the true value  $y$ .
3. **Backward Pass:** Calculate the gradient of the loss with respect to each parameter using the chain rule of calculus:

$$\frac{\partial L}{\partial \mathbf{W}^{(2)}} = \frac{\partial L}{\partial \hat{\mathbf{y}}} \cdot \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{W}^{(2)}} = \boldsymbol{\delta}^{(2)} \mathbf{h}^T \quad (3.9)$$

$$\frac{\partial L}{\partial \mathbf{b}^{(2)}} = \boldsymbol{\delta}^{(2)} \quad (3.10)$$

$$\frac{\partial L}{\partial \mathbf{W}^{(1)}} = \frac{\partial L}{\partial \mathbf{h}} \cdot \frac{\partial \mathbf{h}}{\partial \mathbf{W}^{(1)}} = \boldsymbol{\delta}^{(1)} \mathbf{x}^T \quad (3.11)$$

$$\frac{\partial L}{\partial \mathbf{b}^{(1)}} = \boldsymbol{\delta}^{(1)} \quad (3.12)$$

where  $\boldsymbol{\delta}^{(2)} = \frac{\partial L}{\partial \hat{\mathbf{y}}} \cdot g'(\mathbf{W}^{(2)}\mathbf{h} + \mathbf{b}^{(2)})$  and  $\boldsymbol{\delta}^{(1)} = (\mathbf{W}^{(2)})^T \boldsymbol{\delta}^{(2)} \cdot f'(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)})$  represent the error signals at each layer.

4. **Parameter Update:** Update all parameters using gradient descent:

$$\mathbf{W}^{(2)} \leftarrow \mathbf{W}^{(2)} - \eta \frac{\partial L}{\partial \mathbf{W}^{(2)}} \quad (3.13)$$

$$\mathbf{b}^{(2)} \leftarrow \mathbf{b}^{(2)} - \eta \frac{\partial L}{\partial \mathbf{b}^{(2)}} \quad (3.14)$$

$$\mathbf{W}^{(1)} \leftarrow \mathbf{W}^{(1)} - \eta \frac{\partial L}{\partial \mathbf{W}^{(1)}} \quad (3.15)$$

$$\mathbf{b}^{(1)} \leftarrow \mathbf{b}^{(1)} - \eta \frac{\partial L}{\partial \mathbf{b}^{(1)}} \quad (3.16)$$

where  $\eta$  is the learning rate.

The power of backpropagation lies in its computational efficiency. Rather than recomputing gradients for each parameter independently, it reuses intermediate results through the chain rule, significantly reducing computational complexity. This efficiency enables the training of deep neural networks with millions of parameters on large datasets [35].

For complex architectures like CNNs and Transformers, the principle remains the same, though the specific gradient calculations become more intricate due to operations like convolution and self-attention. Modern deep learning frameworks like PyTorch provide automatic differentiation capabilities that implement backpropagation transparently [36], allowing researchers to focus on model architecture rather than gradient calculations.

### 3.2.4 Optimization

Once the gradients of the loss function with respect to the model parameters are computed using backpropagation, optimization algorithms use these gradients to update the parameters and minimize the loss. The most fundamental approach is gradient descent, which updates parameters in the direction of steepest descent of the loss function:

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} L(\theta_t) \quad (3.17)$$

where  $\theta$  represents the model parameters,  $\eta$  is the learning rate, and  $\nabla_{\theta} L(\theta_t)$  is the gradient of the loss function with respect to the parameters.

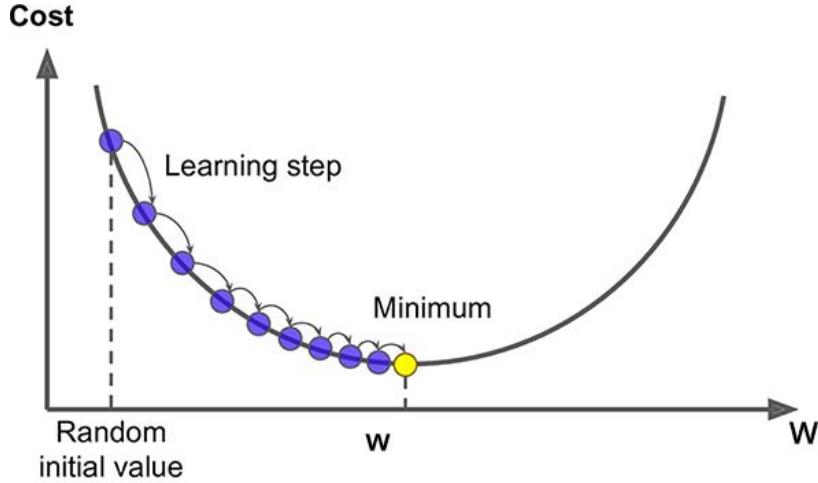


Figure 3.3: Illustration of gradient descent optimization. Starting from a random initial value, each learning step moves the parameters ( $w$ ) in the direction of steepest descent until reaching a minimum of the cost function [37].

In practice, stochastic gradient descent (SGD) and its variants are more commonly used, where gradients are computed on subsets of the data (mini-batches) rather than the entire dataset. This approach is more computationally efficient and often leads to better generalization [38].

For our charged Higgs analysis, we use the Adam (Adaptive Moment Estimation) optimizer, which combines the benefits of two other extensions of SGD—AdaGrad and RMSProp [39]. Adam computes adaptive learning rates for each parameter based on estimates of the first and second moments of the gradients:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (3.18)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (3.19)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (3.20)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (3.21)$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \quad (3.22)$$

where  $g_t$  is the gradient at time  $t$ ,  $m_t$  and  $v_t$  are the first and second moment estimates,  $\beta_1$  and  $\beta_2$  are decay rates, and  $\epsilon$  is a small constant to prevent division by zero.

### 3.2.5 Regularization

A common challenge in neural network training is overfitting, which occurs when the model learns the training data too well, including its noise and peculiarities, resulting in poor generalization to unseen data. Various regularization techniques are employed to prevent overfitting:

**L2 regularization** (weight decay) adds a penalty term to the loss function based on the squared magnitude of the weights:

$$\mathcal{L}_{\text{reg}} = \mathcal{L} + \lambda \sum_i w_i^2 \quad (3.23)$$

where  $\lambda$  is the regularization strength. This discourages the model from learning large weights, promoting simpler models that are less likely to overfit.

**Dropout** is another powerful regularization technique where a random subset of neurons is temporarily deactivated during each training iteration. Dropout forces the network to learn redundant representations and prevents co-adaptation of neurons, acting as an implicit ensemble method [40].

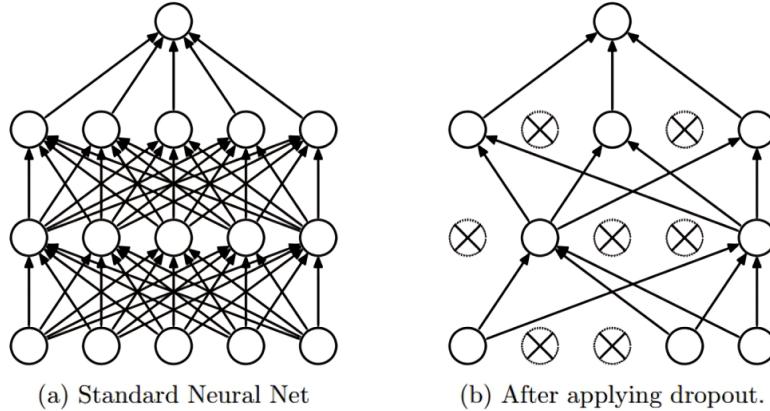


Figure 3.4: Illustration of dropout regularization: (a) Standard Neural Network with all neurons active, and (b) Network after applying dropout with randomly deactivated neurons (marked with X) [41].

**Early stopping** monitors performance on a validation set during training and stops when validation performance begins to degrade, capturing the model at its optimal generalization point before overfitting occurs.

**Batch normalization** normalizes the inputs of each layer to have zero mean and unit variance within each mini-batch:

$$\hat{x} = \frac{x - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad y = \gamma \hat{x} + \beta \quad (3.24)$$

where  $\mu_B$  and  $\sigma_B^2$  are the mean and variance of the mini-batch, and  $\gamma$  and  $\beta$  are learnable parameters. Batch normalization stabilizes and accelerates training by reducing internal covariate shift, while also providing some regularization benefits.

## 3.3 Deep Learning Architectures for Charged Higgs Search

### 3.3.1 Problem Formulation

We formulate the charged Higgs boson search as a binary classification problem for each decay channel (lvbb and qqbb) separately. Given the low-level particle data for an event, the model predicts whether the event contains a true charged Higgs boson decaying through the specific channel.

Each event is represented as a set of particles, with each particle described by a five-dimensional feature vector: the three-momentum components ( $p_x$ ,  $p_y$ ,  $p_z$ ), energy ( $E$ ), and particle type. The particle type is encoded as an integer: -1 for padding (not a real particle), 0 for electron, 1 for muon, 2 for neutrino, 3 for large-radius jet, and 4 for small-radius jet.

For the lvbb channel model, the target variable is 1 if the event has `truth_W_decay_mode = 1` (indicating a leptonic W decay) and 0 otherwise. Similarly, for the qqbb channel model, the target is 1 if `truth_W_decay_mode = 2` (indicating a hadronic W decay) and 0 otherwise.

Each event also has an associated weight that accounts for cross-section differences between processes and detector acceptance effects. These weights are incorporated into the loss function during training to ensure the model learns the correct physics-motivated class balance, as detailed in Chapter 2.

### 3.3.2 Data Preparation for Neural Networks

#### Handling Variable Particle Counts

A fundamental challenge in using particle data for machine learning is that different events contain different numbers of particles, while neural networks typically require fixed-size inputs. This variability is inherent to the physics—different collision events produce different numbers of particles depending on the specific process and energy involved. To address this, we set a maximum number of particles per event (`max_num_particles = 30`), truncate events with more particles (keeping only the first 30), and zero-pad events with fewer particles to reach the maximum. This approach ensures a consistent input size across all events while preserving as much information as possible.

For the particle type feature, we use a special value of -1 for padding entries to distinguish them from real particles. This value choice is significant as it lies outside the range of valid particle types (0-4), ensuring that the model can easily identify and ignore padding.

To ensure that the padding is properly handled during model training, we also create masks based on the padding values. These masks are used in both the CNN and Transformer models to ensure that padding does not contribute to the learned features, preventing the model from learning spurious correlations from non-physical padding values.

## Particle Type Encoding

The particle type is a categorical feature, which needs special handling in machine learning models. For the CNN model, we use the integer encoding directly as one of the input channels. For the Transformer model, we employ an embedding layer to map these integers to learned vector representations. This approach is physically motivated, as it allows the model to learn relationships between different particle types that might not be captured by a simple one-hot encoding [42].

The embedding dimension is set to 16, allowing the model to learn meaningful representations of different particle types that capture their physical relationships. This approach is more powerful than one-hot encoding, as it can learn to place similar particle types (e.g., electrons and muons, which are both leptons) close to each other in the embedding space, reflecting their physical similarities. This learned embedding space may reveal interesting physical insights about the relationships between different particle types in the context of charged Higgs searches.

## Data Shape and Structure

After preprocessing, the data has a specific structure designed for input to our neural network models. The input features form a 3D array of shape  $(\text{num\_events}, \text{num\_features}, \text{max\_num\_particles}) = (\text{num\_events}, 5, 30)$ , where the features are  $p_x$ ,  $p_y$ ,  $p_z$ ,  $E$ , and particle type. The labels are a 1D array of shape  $(\text{num\_events},)$  with binary values indicating signal (1) or background (0), and the event weights form another 1D array of shape  $(\text{num\_events},)$  containing the physics-motivated weights.

This structure is suitable for input to the convolutional layers of the CNN model. For the Transformer model, the data is transposed to  $(\text{num\_events}, \text{max\_num\_particles}, \text{num\_features})$  to align with the expected input format, where each particle is treated as a token in the sequence. This transformation preserves all the physical information while adapting to the architectural requirements of each model type.

## Training, Validation, and Test Strategy

As mentioned in Chapter 2, we split the preprocessed data into training (70%), validation (15%), and test (15%) sets with stratification to maintain consistent class distributions. Beyond the basic splitting strategy, several considerations were made specifically for the machine learning workflow:

- **Early Stopping:** The validation set is monitored during training to determine when to stop the optimization process. Training terminates when the validation loss does not improve for a specified number of epochs (patience = 5), preventing overfitting.
- **Mass Point Representation:** Special care was taken to ensure all charged Higgs mass points were proportionally represented in all data splits, enabling the model to learn from and generalize across the entire mass range.
- **Consistent Evaluation:** All models were evaluated on the same test set to enable fair comparisons between different architectural approaches.

This comprehensive data preparation strategy addresses the unique challenges of applying deep learning to particle physics data, creating a robust foundation for developing effective charged Higgs identification models.

### 3.3.3 Convolutional Neural Network Architecture

Convolutional Neural Networks (CNNs) are specialized neural network architectures originally designed for image processing but adaptable to various data types with grid-like topology, including sequences of particles [43]. For our charged Higgs identification task, we implement a CNN model with an attention mechanism to enhance its discriminative power.

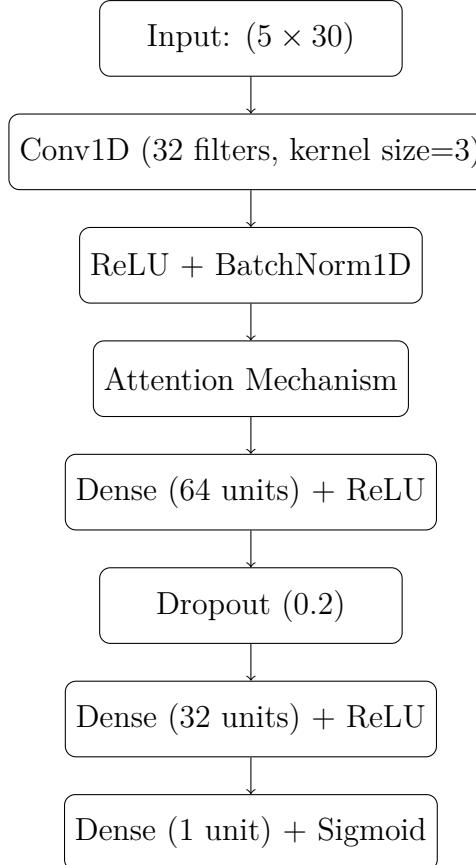


Figure 3.5: Architecture of the Convolutional Neural Network model with attention mechanism used for charged Higgs identification. The model processes particle sequences through convolutional layers followed by an attention mechanism that identifies the most relevant particles.

The key components of our CNN architecture are:

#### Convolutional Layer (Conv1D)

The core building block of our CNN is a one-dimensional convolutional layer that scans through the particle sequence. Conceptually, this layer performs a sliding window operation, where a set of learnable filters (also called kernels) moves across the particle sequence, examining a few particles at a time [43]. Each filter looks for specific patterns—such as certain momentum correlations or energy distributions—that might be indicative of signal events.

In our implementation, we use 32 filters with a kernel size of 3, meaning each filter examines 3 adjacent particles at a time. The filters are applied with padding to ensure the output maintains the same length as the input. The primary advantage

of this approach is that the same filter is applied across the entire sequence, enabling the network to detect important patterns regardless of where they appear among the particles.

### Batch Normalization and ReLU

After the convolutional operation, the outputs are normalized using batch normalization and then passed through a ReLU activation function. As explained in Section 3.2.6, batch normalization stabilizes training by normalizing the layer outputs, while ReLU introduces non-linearity necessary for the network to learn complex patterns.

### Attention Mechanism

One of the most innovative aspects of our CNN architecture is the incorporation of an attention mechanism. The fundamental challenge in particle physics data is that not all particles are equally important for determining whether an event contains a charged Higgs boson. Some particles are direct decay products of the Higgs or W boson, while others may be less relevant background particles.

The attention mechanism addresses this challenge by learning to assign importance weights to different particles [44]. Conceptually, it works by:

1. Examining the features of each particle extracted by the convolutional layer
2. Computing an "importance score" for each particle based on these features
3. Converting these scores into a set of weights (summing to 1) using a softmax function
4. Creating a weighted sum of particle features, emphasizing the most important particles

This approach allows the model to focus on the most discriminative particles—likely those originating from Higgs or W boson decays—while downplaying less informative background particles.

### Fully Connected Layers

The output from the attention mechanism is a fixed-size vector representing the entire event. This vector is processed through a series of fully connected layers (essentially, Neural Network) with ReLU activations and dropout regularization. These layers progressively transform the high-level features extracted by the previous layers into a prediction about whether the event contains a charged Higgs boson. The final layer uses a sigmoid activation function to output a probability between 0 and 1.

In summary, our CNN architecture combines local pattern detection (through convolutions) with adaptive focus on important particles (through attention) to effectively identify charged Higgs events from low-level particle data. This approach allows the model to learn directly from raw particle features without requiring manual high-level feature engineering.

### 3.3.4 Transformer Architecture

Transformer models represent a fundamentally different approach to sequence processing compared to CNNs. While CNNs excel at identifying local patterns through sliding window operations, Transformers are designed to capture relationships between all pairs of elements in a sequence, regardless of their distance from each other [42]. This global modeling capability makes Transformers particularly interesting for particle physics, where correlations between particles may extend beyond local neighborhoods.

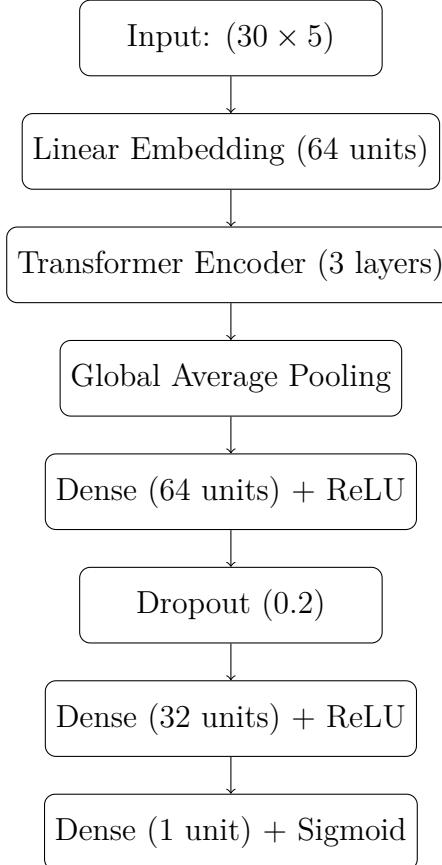


Figure 3.6: Architecture of the Transformer model used for charged Higgs identification. Each particle is treated as a token in a sequence, enabling the self-attention mechanism to capture global relationships between particles regardless of their positions.

The key components of our Transformer architecture are:

#### Embedding Layer

Unlike the CNN, which works directly with the raw 5-dimensional particle features, the Transformer first projects each particle's features into a higher-dimensional space (64 dimensions in our implementation) through a linear transformation. This embedding step enriches the representation, allowing the model to capture more complex patterns in subsequent layers.

#### Self-Attention Mechanism

The cornerstone of the Transformer architecture is the self-attention mechanism, which allows each particle to attend to all other particles in the event. Conceptually, self-

attention can be understood as a sophisticated way of modeling relationships between particles:

1. For each particle, the model computes three vectors: a Query vector (representing what the particle is "looking for"), a Key vector (representing what the particle "offers" to others), and a Value vector (representing the information the particle contains).
2. The model computes compatibility scores between each particle's Query and all particles' Keys, essentially asking: "How relevant is particle j to particle i?"
3. These compatibility scores are normalized using a softmax function to create attention weights.
4. Each particle's output is computed as a weighted sum of all particles' Value vectors, where the weights are determined by the attention weights [42].

This process allows each particle to gather information from all other particles based on learned relevance, regardless of their distance in the sequence. For charged Higgs identification, this means the model can directly capture relationships between decay products even if they're widely separated in the detector.

Our implementation uses Multi-Head Attention with 4 heads, which allows the model to attend to different types of relationships simultaneously. Each head can focus on different aspects of the physics—for example, one head might focus on momentum correlations while another focuses on angular relationships.

### **Transformer Encoder Layers**

Our model stacks 3 Transformer encoder layers, each containing a self-attention mechanism followed by a feed-forward neural network, with layer normalization and residual connections around each sublayer. This stacking allows the model to build increasingly abstract representations of the event, progressively refining its understanding of the particle relationships.

It's worth noting that unlike standard Transformer implementations, we intentionally omitted positional encoding from our model. Positional encoding is typically used in Transformers to provide information about the sequential order of elements [42]. However, for our particle physics application, the absolute ordering of particles in our data representation is arbitrary and not physically meaningful. What matters are the physical properties and relationships between particles, not their position in the input sequence. This design choice reflects the unique nature of particle physics data compared to naturally sequential data like text or time series.

### **Global Average Pooling**

After processing through the Transformer encoder layers, each particle has an associated feature vector enriched with information about its relationships to other particles. To create a single representation for the entire event, we use Global Average Pooling, which simply takes the average of all particle feature vectors (excluding padding particles). This aggregation step transforms the variable-length sequence of particle representations into a fixed-size event representation suitable for classification [45].

### Classification Head

Similar to the CNN architecture, the pooled event representation is passed through a series of fully connected layers with ReLU activations and dropout regularization, resulting in a sigmoid output that represents the probability of the event containing a charged Higgs boson.

The Transformer architecture’s global modeling capability offers a complementary approach to the CNN’s local pattern detection. While CNNs focus on identifying specific local patterns that might indicate Higgs or W boson decay products, Transformers model the entire event holistically, potentially capturing complex correlations between all particles regardless of their proximity. This global perspective might be particularly valuable for understanding the overall event topology characteristic of charged Higgs production and decay.

## 3.4 Training Procedure

The training procedure incorporates several techniques to ensure effective learning and generalization. We carefully tuned the training parameters for each architecture separately to account for their different characteristics:

- **Loss Function:** Both models use weighted binary cross-entropy loss, incorporating event weights to account for physics-based importance. For the lvbb channel, we applied a class balancing factor that equalizes the total weighted contributions of signal and background events, while for the qqbb channel, we used class weights inversely proportional to class frequencies as described in section 2.4.3.
- **Optimizer:** Both models use the AdamW optimizer with the following specific configurations:
  - CNN model: Initial learning rate =  $1 \times 10^{-3}$
  - Transformer model: Initial learning rate =  $5 \times 10^{-5}$  for lvbb and  $2 \times 10^{-5}$  for qqbb

The significantly lower learning rate for the Transformer model reflects its greater depth and complexity, which requires more careful optimization to prevent divergence during training.

- **Batch Size:** 64 events per batch for both models, as indicated in the `prepare_data_loaders` function. This batch size provides a good balance between computational efficiency and optimization stability.
- **Maximum Epochs:** 20 for CNN models and 15-20 for Transformer models (15 for lvbb, 20 for qqbb).
- **Early Stopping:** For both models, training stops if validation loss does not improve for 5 consecutive epochs. This prevents overfitting by capturing the model at its optimal generalization point.
- **Learning Rate Scheduling:** We implemented a learning rate scheduler with warmup to improve training stability. The warmup period consists of 20% of the

total training steps, during which the learning rate gradually increases from 0 to the initial value. After the warmup period, the learning rate remains constant.

- **Gradient Clipping:** Maximum gradient norm of 1.0 for both models to prevent exploding gradients, which is particularly important for the Transformer model with its deeper architecture.
- **Label Smoothing:** For the Transformer model, we applied a label smoothing factor of 0.1 for both channels to prevent the model from becoming overconfident in its predictions and improve generalization.
- **Regularization:** Several techniques are applied to prevent overfitting:
  - **Dropout:** Applied with a rate of 0.2 in the fully connected layers of both models. In the Transformer model, additional dropout is applied in the encoder layers.
  - **Batch Normalization:** Applied in the CNN model after the convolutional layer to stabilize training and accelerate convergence.
  - **Weight Decay:** L2 regularization with a factor of 0.01 applied through the AdamW optimizer for both models, penalizing large weights to encourage simpler models.

All models were implemented using PyTorch, a flexible deep learning framework that provides efficient tensor computations and automatic differentiation. The implementation includes:

- A custom PyTorch `ParticleDataset` class to handle the particle physics data, including event weights
- For the Transformer model, padding masks to ensure that padding particles do not contribute to the learned features, which is particularly important for the self-attention mechanism
- A custom training loop in `train_model` that incorporates event weights into the loss calculation and implements the early stopping and learning rate scheduling described above

Training was performed locally on a MacBook Pro with Apple M4 Max chip, 36GB of memory. The CNN model took approximately 10-15 minutes to train per channel, while the Transformer model required around 45 minutes - 1 hour and 20 minutes per channel. The significant difference in training time reflects the greater computational complexity of the Transformer architecture, particularly its self-attention mechanism which has quadratic complexity with respect to the sequence length.

## 3.5 Performance Evaluation Metrics

To evaluate the performance of our models, we use a combination of standard machine learning metrics and physics-motivated measures, all calculated with appropriate event weighting to reflect the true physics processes.

### 3.5.1 Standard Classification Metrics

- **Accuracy:** The proportion of correctly classified events:

$$\text{Accuracy} = \frac{\sum_{i=1}^N w_i \cdot I(\hat{y}_i = y_i)}{\sum_{i=1}^N w_i} \quad (3.25)$$

- **Precision:** The proportion of predicted signal events that are actually signal:

$$\text{Precision} = \frac{\sum_{i=1}^N w_i \cdot I(\hat{y}_i = 1 \text{ and } y_i = 1)}{\sum_{i=1}^N w_i \cdot I(\hat{y}_i = 1)} \quad (3.26)$$

- **Recall (Signal Efficiency):** The proportion of actual signal events that are correctly predicted:

$$\text{Recall} = \frac{\sum_{i=1}^N w_i \cdot I(\hat{y}_i = 1 \text{ and } y_i = 1)}{\sum_{i=1}^N w_i \cdot I(y_i = 1)} \quad (3.27)$$

- **F1-Score:** The harmonic mean of precision and recall:

$$\text{F1-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.28)$$

where  $I(\text{condition})$  is the indicator function that returns 1 if the condition is true and 0 otherwise, and  $\hat{y}_i$  is the predicted label after applying a threshold of 0.5 to the model's output probability.

### 3.5.2 ROC Curve and AUC

The Receiver Operating Characteristic (ROC) curve plots the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The ROC curve is particularly valuable in HEP searches as it illustrates the trade-off between signal efficiency and background rejection across all possible classification thresholds.

The true positive rate and false positive rate are calculated as:

$$\text{TPR}(\theta) = \frac{\sum_{i=1}^N w_i \cdot I(p_i \geq \theta \text{ and } y_i = 1)}{\sum_{i=1}^N w_i \cdot I(y_i = 1)} \quad (3.29)$$

$$\text{FPR}(\theta) = \frac{\sum_{i=1}^N w_i \cdot I(p_i \geq \theta \text{ and } y_i = 0)}{\sum_{i=1}^N w_i \cdot I(y_i = 0)} \quad (3.30)$$

where  $p_i$  is the predicted probability for sample  $i$ ,  $\theta$  is the classification threshold,  $w_i$  is the event weight, and  $I(\text{condition})$  is the indicator function.

The Area Under the Curve (AUC) provides a single scalar value representing the overall performance of the classifier across all possible thresholds:

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}^{-1}(x)) dx \quad (3.31)$$

The AUC has an important statistical interpretation: it represents the probability that a randomly chosen signal event receives a higher score than a randomly chosen background event. An AUC of 0.5 indicates performance no better than random classification, while an AUC of 1.0 represents perfect classification. In practice, AUC values above 0.8 typically indicate good discrimination power in HEP applications [46].

For charged Higgs searches, the ROC curve and AUC are particularly useful because:

- They are insensitive to the specific choice of threshold, providing an overall assessment of discriminative power
- They account for event weights, reflecting the true physics processes
- They allow for direct comparison between different models and approaches
- They provide a framework for selecting optimal working points based on specific analysis needs (e.g., prioritizing signal efficiency or background rejection)

### 3.5.3 Physics-Motivated Metrics

In addition to standard machine learning metrics, we also consider metrics that are particularly relevant for high-energy physics analyses:

- **Signal Efficiency:** The fraction of signal events that are correctly classified as signal, equivalent to recall. In HEP, maintaining high signal efficiency is often prioritized to ensure potential discoveries are not missed.
- **Background Rejection:** The fraction of background events that are correctly classified as background, calculated as  $1 - \text{FPR}$ . This metric is crucial for suppressing the typically overwhelming background in HEP searches.

These metrics provide a comprehensive evaluation of the models' performance, both from a machine learning perspective and in terms of their potential impact on charged Higgs searches at the LHC. The results of this evaluation are presented in detail in Chapter 4.

# Chapter 4

## Results and Performance Evaluation

### 4.1 Model Training Assessment

The CNN model applied to the lvbb channel shows variable training dynamics with fluctuations in validation loss. Despite this volatility, the overall trend was clearly downward, with this model training for 20 epochs and its loss functions hitting a plateau, indicating that the model has reached its optimal performance.

In contrast, the Transformer for lvbb shows steady convergence with the validation loss closely tracking training loss, indicating good generalization without overfitting. The smoother nature of learning in the Transformer model indicates that it may encounter local minima much less frequently than the CNN, due to difference in optimization dynamics of these 2 models.

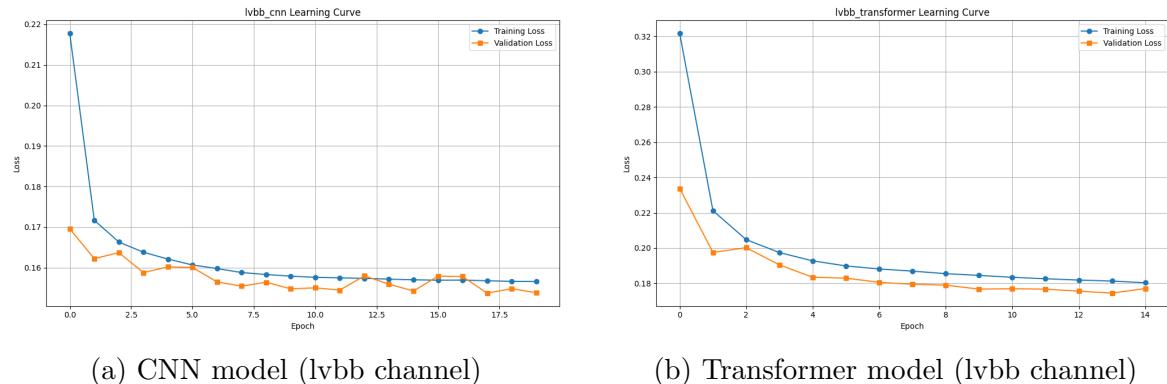


Figure 4.1: Learning curves for the lvbb channel models showing training loss (blue) and validation loss (orange) as functions of training epoch.

For the CNN model in qqbb channel, both training and validation losses decreased smoothly with minimal separation between them, indicating exceptional generalization capability. The model trained for 20 epochs, after which loss plateaued. The Transformer for qqbb showed training characteristics similar to its lvbb counterpart—more difference between loss functions than the CNN but with satisfactory convergence. Both models trained for 20 epochs.

The CNN model in the qqbb channel demonstrated the most stable training dynamics across all channels, with nearly ideal convergence behavior. Transformers exhibited

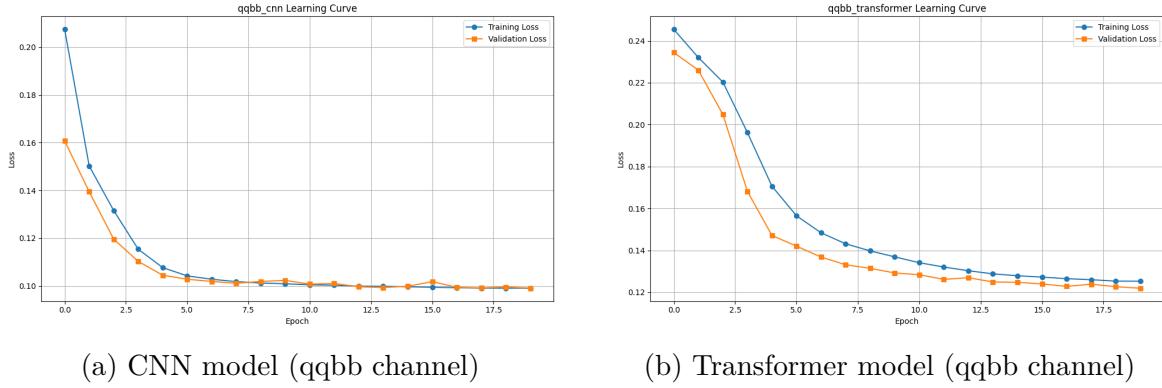


Figure 4.2: Learning curves for the qqbb channel models showing training loss (blue) and validation loss (orange) as functions of training epoch.

a healthy training procedure with consistent progression, showing only a modest gap between training and validation losses—reflecting their more sophisticated modeling capabilities. This controlled learning process enabled Transformers to achieve superior performance in the lvbb channel by effectively capturing complex particle relationships while maintaining good generalization properties.

## 4.2 Overall Classification Performance

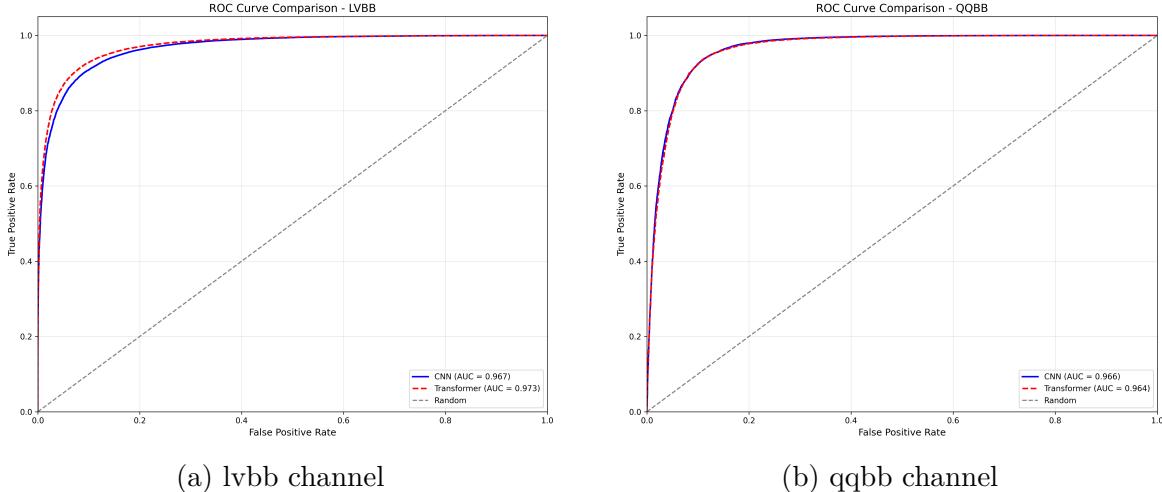


Figure 4.3: ROC curves for CNN (blue) and Transformer (red) models on both decay channels. The AUC values are 0.967 (CNN) and 0.973 (Transformer) for lvbb, and 0.966 (CNN) and 0.964 (Transformer) for qqbb.

All models achieve exceptional AUC values exceeding 0.96, confirming that both neural network architectures effectively learn meaningful patterns from low-level particle information. This performance substantially surpasses typical cut-based analyses.

In the leptonic channel, the Transformer model shows a modest but consistent advantage over the CNN (AUC 0.973 vs. 0.967), particularly in high-purity regions. For the hadronic channel, performance differences narrow considerably (AUC 0.966 for

CNN vs. 0.964 for Transformer), with the CNN performing better in high-efficiency regions.

Comparing across channels, the best models achieve remarkably similar overall discrimination capability despite substantially different decay topologies and background compositions, suggesting both channels offer viable paths for charged Higgs searches.

## 4.3 Detailed Performance Analysis

### 4.3.1 Confusion Matrices

In the leptonic (lvbb) channel, the Transformer model achieves slightly higher performance with 92.32% signal efficiency and 90.73% background rejection, compared to the CNN model's 91.29% signal efficiency and 89.52% background rejection. While these signal efficiencies are lower than those seen in the qqbb channel, the background rejection rates are notably higher (approx. 90% vs. 85% for qqbb). The Transformer's modest advantage in both efficiency and rejection aligns with its higher AUC value (0.973 vs. 0.967 for the CNN), suggesting its architecture is particularly well-suited for identifying the complex patterns in leptonic decays.

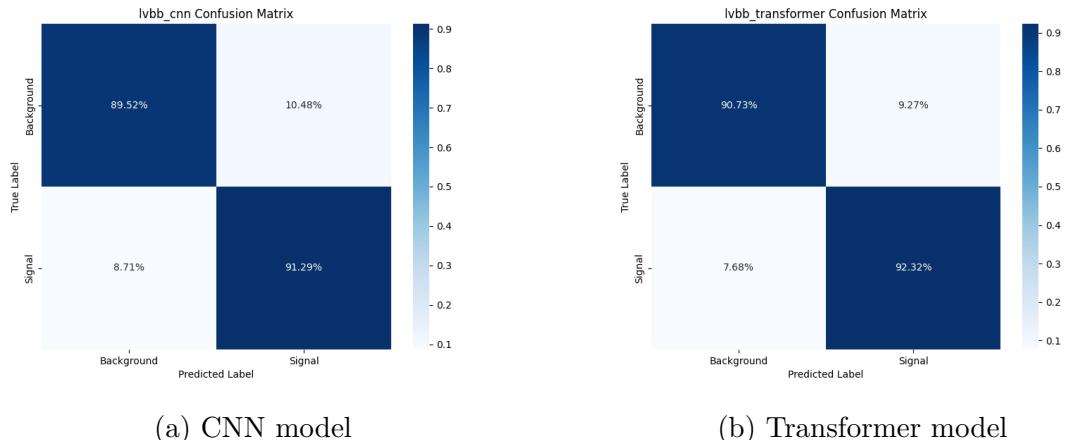


Figure 4.4: Confusion matrices for the lvbb channel models. Values represent the percentage of events in each category after applying a classification threshold of 0.5.

In the hadronic channel, both architectures achieve significantly higher signal efficiencies—96.12% for CNN and 96.13% for Transformer. This comes with lower background rejection rates (85.33% for CNN, 84.99% for Transformer) compared to the lvbb channel. The nearly identical signal efficiencies despite slightly different AUCs suggest both architectures effectively capture the key features of hadronic charged Higgs decays.

A striking pattern emerges when comparing across channels: both architectures achieve significantly higher signal efficiency in qqbb (96% vs. 91-92% for lvbb) but better background rejection in lvbb (90% vs. 85% for qqbb). This consistent pattern reveals complementary strengths of the two decay channels that could be exploited in a combined search strategy.

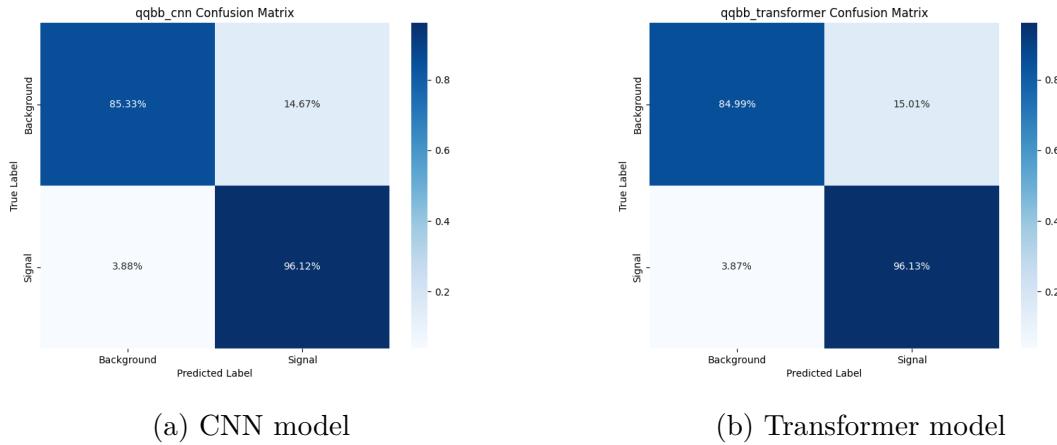


Figure 4.5: Confusion matrices for the qqbb channel models. Values represent the percentage of events in each category after applying a classification threshold of 0.5.

### 4.3.2 Prediction Score Distributions

For the lvbb channel, the CNN produces a signal distribution strongly peaked near 1.0, while the background distribution concentrates near 0.0 with a moderate tail extending toward higher scores. The Transformer generates even more sharply peaked distributions for both classes, explaining its superior performance metrics. The reduced overlap between signal and background distributions for the Transformer confirms its more confident and accurate classifications in the leptonic channel.

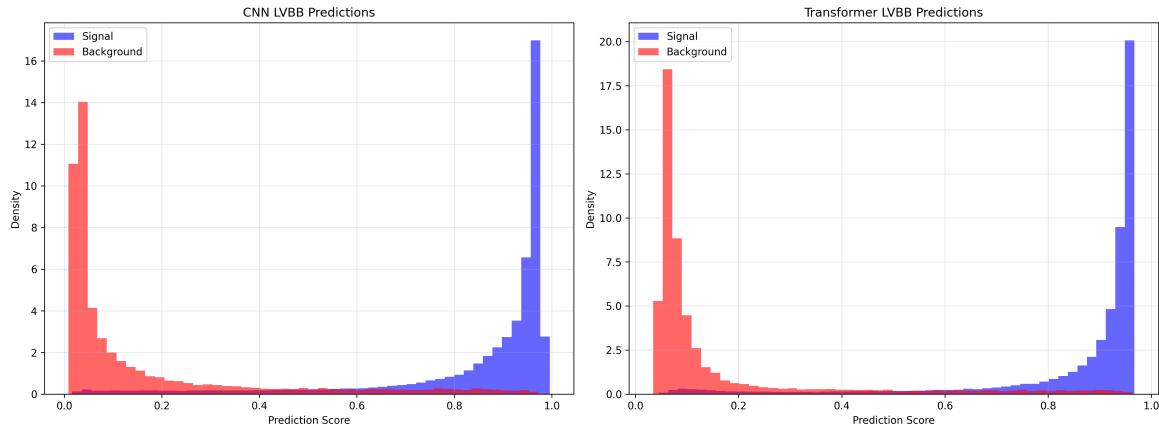


Figure 4.6: Prediction score distributions for signal (blue) and background (red) events in the lvbb channel, comparing CNN (left) and Transformer (right) models. The vertical dashed line indicates the default classification threshold of 0.5.

In the qqbb channel, both architectures produce remarkably sharp signal distributions with minimal spread, explaining their high signal efficiencies. The background distributions show more pronounced tails than in lvbb, consistent with the lower background rejection rates. The CNN's background distribution displays a slightly thinner tail, aligning with its marginally better background rejection.

Comparing across channels, both architectures produce more sharply peaked signal distributions for qqbb than for lvbb, confirming that neural networks identify hadronic charged Higgs decays with greater confidence. The more pronounced background tails

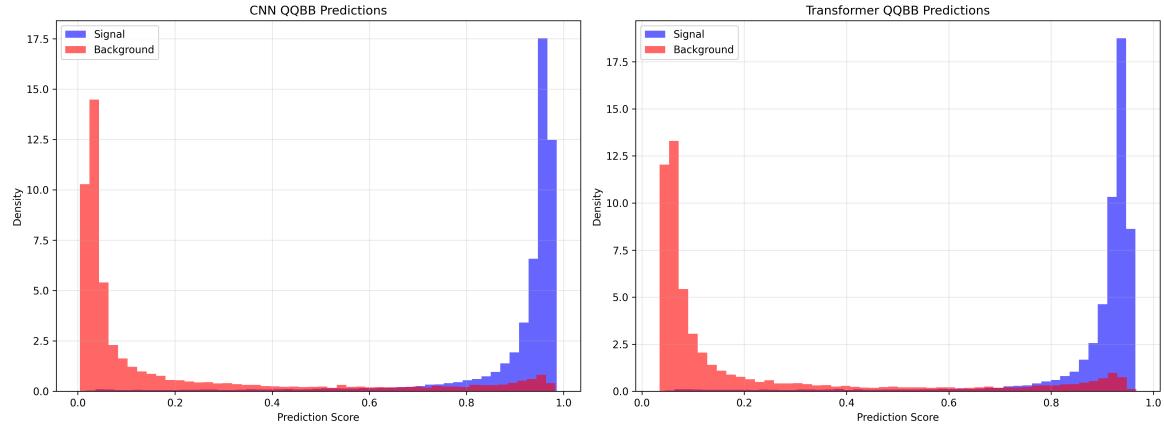


Figure 4.7: Prediction score distributions for signal (blue) and background (red) events in the qqbb channel, comparing CNN (left) and Transformer (right) models. The vertical dashed line indicates the default classification threshold of 0.5.

in qqbb suggest greater similarity between some background processes and hadronic signal events, explaining the lower background rejection rates in this channel.

## 4.4 Comparative Analysis of Architectures and Channels

Table 4.1 presents a detailed quantitative comparison of model performance across architectures and decay channels.

Metric	LVBB CNN	LVBB Trans.	QQBB CNN	QQBB Trans.
Accuracy	0.9041	0.9153	<b>0.9310</b>	0.9301
ROC AUC	0.9672	<b>0.9729</b>	0.9656	0.9643
Precision	0.8974	0.9091	<b>0.9440</b>	0.9428
Recall	0.9129	0.9232	0.9612	<b>0.9613</b>
F1 Score	0.9051	0.9161	<b>0.9525</b>	0.9520

Table 4.1: Performance comparison across models and channels. Best results for each metric are highlighted in bold.

This comparative analysis reveals several key patterns:

- The hadronic channel demonstrates superior performance on most metrics, suggesting this decay signature may be more consistently identifiable with low-level particle information.
- The Transformer consistently outperforms the CNN in the leptonic channel, while the CNN slightly edges out in the hadronic channel. These preferences align with the underlying physics—global relationships in leptonic decays benefit from attention mechanisms, while localized jet patterns in hadronic decays suit convolutional operations.

- Both models achieve notably higher recall in the qqbb channel compared to lvbb, indicating substantially better signal identification capability for hadronic decays regardless of architecture choice.
- The consistently higher performance in the qqbb channel is particularly noteworthy, as this channel typically presents greater challenges in conventional cut-based analyses due to higher background contamination.

These findings have important implications for search strategies. The Transformer’s superior global modeling capability proves valuable for lvbb events, which involve missing energy and complex correlations between particles. The CNN’s strong performance in qqbb aligns with the localized energy patterns and jet substructure characteristic of hadronic decays.

The qqbb channel’s higher signal efficiency makes it more sensitive for discovering charged Higgs bosons, especially with low production cross-sections. Conversely, the lvbb channel’s superior background rejection could be advantageous for setting stringent limits where background contamination is the limiting factor.

## 4.5 Physics Interpretation and Summary

The exceptional performance of neural networks trained on low-level particle information ( $AUCs > 0.96$ ) validates that deep learning can extract meaningful patterns directly from basic kinematic properties without engineered features. The 91-96% signal efficiencies dramatically exceed typical cut-based approaches (40-60%), suggesting these models identify subtle patterns that conventional analyses might miss.

Based on the achieved signal efficiencies and background rejections, these approaches could potentially improve search sensitivity by factors of 2-3 compared to traditional methods. Even in high-purity regions of the ROC curves (false positive rates  $< 1\%$ ), the models maintain signal efficiencies of 50-70%, enabling searches that require strict background control.

The complementary strengths of the two channels—qqbb’s high signal efficiency and lvbb’s superior background rejection—suggest that a combined analysis would maximize discovery potential across the charged Higgs mass range. This would be particularly valuable for intermediate masses (1000-2000 GeV) where both channels show strong performance.

In conclusion, this chapter has demonstrated that both CNN and Transformer architectures effectively learn to identify charged Higgs events from low-level particle information. The comparative analysis reveals channel-specific architectural strengths that align with the underlying physics of different decay modes. The performance achieved by these models suggests potential for substantial improvements in charged Higgs search sensitivity. The following chapter explores limitations of the current approach and directions for future work in this promising intersection of particle physics and machine learning.

# Chapter 5

## Conclusion and Future Work

### 5.1 Summary of Findings

This dissertation has successfully demonstrated the application of deep learning techniques to low-level particle information for charged Higgs boson identification, addressing the research objectives outlined in Chapter 1. The work has made several significant contributions to the field of particle physics analysis and machine learning applications in high-energy physics.

The development of neural network models capable of learning directly from basic kinematic properties of particles has proven remarkably effective, achieving high discrimination between signal and background events without relying on manually engineered high-level features. This approach could represent a shift from traditional cut-based analyses, allowing the models to potentially discover subtle patterns that might otherwise be overlooked in conventional approaches.

Our comparative analysis revealed important architectural preferences aligned with the underlying physics of different decay modes. The Transformer architecture demonstrated superior performance in the leptonic channel (lvbb), where the event topology involves complex global relationships between visible particles and missing energy, achieving an impressive AUC of 0.973, accuracy of 91.53%, and F1 score of 91.61%. Conversely, the CNN architecture showed strength in the hadronic channel (qqbb), with an AUC of 0.966, accuracy of 93.10%, and F1 score of 95.25%. These performance metrics highlight how each architecture's inherent strengths align with the physical characteristics of different decay modes.

Perhaps most significantly, we identified complementary strengths between the decay channels that could be leveraged in future searches. The hadronic channel enables substantially higher signal efficiency (96.12-96.13% compared to 91.29-92.32% for lvbb), making it particularly valuable for discovery scenarios where maximizing signal acceptance is crucial. Meanwhile, the leptonic channel provides superior background rejection (approximately 90% compared to 85% for qqbb), offering advantages for setting stringent limits where background contamination is the limiting factor.

The remarkably high recall values achieved by both models in the qqbb channel (96.12% for CNN and 96.13% for Transformer) suggest that hadronic decays of charged Higgs bosons present distinctive signatures that neural networks can identify with exceptional effectiveness, despite the traditionally challenging hadronic environment. Similarly, the precision values in the qqbb channel (94.40% for CNN and 94.28% for Transformer) significantly outperform those in the lvbb channel (89.74% for CNN

and 90.91% for Transformer), indicating fewer false positives in the hadronic model predictions.

These findings collectively suggest that the application of deep learning to low-level particle analysis has significant potential to enhance the sensitivity of charged Higgs searches at the LHC, potentially enabling exploration of previously inaccessible regions of parameter space in beyond the Standard Model theories with extended Higgs sectors.

## 5.2 Limitations and Challenges

Despite the promising results achieved, this study faced several important limitations that must be acknowledged when interpreting the findings and considering their applicability to experimental searches.

A fundamental limitation is the reliance on simulated data rather than real detector measurements from the ATLAS experiment. While simulation incorporates sophisticated models of detector response, it cannot perfectly reproduce all aspects of real collision events, potentially creating a domain gap that could affect model performance when applied to actual experimental data.

The analysis did not incorporate a comprehensive treatment of systematic uncertainties, including detector effects and theoretical uncertainties in the simulation. These uncertainties could significantly impact the reported performance metrics, particularly in regions of parameter space where signal and background predictions are sensitive to modeling assumptions.

Our study focused exclusively on the  $H^+ \rightarrow W^+ h \rightarrow W^+ b\bar{b}$  decay channel, while charged Higgs bosons could also be discovered through other decay modes such as  $H^+ \rightarrow tb$  or  $H^+ \rightarrow \tau^+ \nu_\tau$ . The effectiveness of the low-level particle approach for these alternative channels remains unexplored.

The models were trained and evaluated on a limited set of mass points (800-3000 GeV), with evaluation focused primarily on aggregate performance. Mass-dependent performance variations were not comprehensively investigated, limiting our understanding of how these approaches might perform across the full mass range of interest.

Finally, the dataset employed a uniform distribution of charged Higgs mass points rather than the natural production cross-section, which would decrease significantly with increasing mass. This affects how the reported aggregate performance would translate to an actual experimental search covering the full mass range.

## 5.3 Future Directions

Building on the findings and addressing the limitations of this work, several promising avenues for future research emerge:

### Application to real detector data

A critical next step would be applying these models to actual ATLAS experimental data, potentially through a control region study where known Standard Model processes could be used to validate the approach. This would require developing robust methods to handle the domain gap between simulation and real data, possibly through domain adaptation techniques or adversarial training approaches.

### Incorporation of systematic uncertainties

Future work should integrate a comprehensive treatment of systematic uncertainties into the analysis framework. This could involve developing uncertainty-aware training procedures that account for variations in detector response and theoretical predictions, potentially improving model robustness and enabling more realistic sensitivity estimates.

### Extension to additional decay channels

The low-level particle approach could be extended to other charged Higgs decay channels such as  $H^+ \rightarrow tb$  and  $H^+ \rightarrow \tau^+ \nu_\tau$ . These channels present different experimental signatures and background compositions, potentially offering complementary sensitivity to different regions of parameter space.

### Architectural innovations

Several promising model improvements could be explored, including:

- Hybrid architectures combining the strengths of CNNs (local pattern recognition) and Transformers (global relationship modeling)
- Learned particle embeddings instead of integer encodings to better represent physical relationships between particle types
- Mass-specific models optimized for different charged Higgs mass ranges
- Ensemble methods combining multiple architectures to improve performance

### Advanced interpretability methods

Developing more sophisticated techniques for interpreting what the neural networks have learned could provide valuable physics insights beyond classification performance. Techniques such as attention visualization, feature attribution methods, or physics-informed latent space analysis could reveal which physical properties drive model decisions.

### End-to-end optimization frameworks

Future research could explore frameworks that directly optimize for discovery significance rather than classification metrics, better aligning the machine learning objective with the physics goal. This might involve custom loss functions incorporating the statistical procedures used in actual experimental searches.

### Combined channel analysis

Developing methodologies for statistically combining the results from different decay channels could maximize overall sensitivity to charged Higgs bosons. This would leverage the complementary strengths identified in this dissertation—the high signal efficiency of qqbb (96.12–96.13%) and the superior background rejection of lvbb (approximately 90%)—to create a more powerful unified search strategy.

## 5.4 Broader Impact

The methods developed in this dissertation have potential applications beyond charged Higgs searches. The approach of using low-level particle information with deep learning could be adapted to other beyond the Standard Model searches, including supersymmetry, dark matter candidates, and exotic resonances. The architectural insights regarding the suitability of different neural network designs for specific event topologies may inform model selection across the broader landscape of particle physics analyses.

Additionally, the interpretability framework developed to understand model decisions contributes to the growing body of knowledge about making deep learning applications in physics more transparent and physically meaningful. This addresses one of the key challenges in adopting advanced machine learning methods in scientific domains where understanding the underlying reasoning is as important as achieving high performance.

## 5.5 Final Remarks

This dissertation has demonstrated the significant potential of low-level particle analysis with deep learning techniques to enhance the sensitivity of charged Higgs boson searches at the LHC. The excellent performance metrics achieved—with AUCs exceeding 0.96 across all models, signal efficiencies ranging from 91.29% to 96.13%, and F1 scores between 90.51% and 95.25%—highlight the effectiveness of this approach.

By allowing neural networks to learn directly from basic kinematic properties rather than engineered features, we enable more comprehensive exploitation of the information contained in collision events. The complementary strengths of different architectures and decay channels identified in this work provide a foundation for more effective search strategies.

As experimental searches continue to probe the boundaries of the Standard Model, such approaches that maximize the discovery potential of existing data will play an increasingly important role in our quest to understand the fundamental structure of matter and energy.

# Bibliography

- [1] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett. B* **716**, 1 (2012)
- [2] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett. B* **716**, 30 (2012)
- [3] W.N. Cottingham and D.A. Greenwood, “An Introduction to the Standard Model of Particle Physics,” Cambridge University Press, 2nd Edition (2007)
- [4] M. Kobayashi et al., “Recent progress in charged Higgs boson searches at the LHC,” *Int. J. Mod. Phys. A* **36**, 2130006 (2021)
- [5] A. Howard, “Unsolved Problems in the Standard Model and Beyond,” *Rep. Prog. Phys.* **83**, 086201 (2020)
- [6] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher and J. P. Silva, “Theory and phenomenology of two-Higgs-doublet models,” *Phys. Rept.* **516**, 1 (2012)
- [7] J. F. Gunion, H. E. Haber, G. L. Kane and S. Dawson, “The Higgs Hunter’s Guide,” *Front. Phys.* **80**, 1 (2000)
- [8] A. G. Akeroyd et al., “Prospects for charged Higgs searches at the LHC,” *Eur. Phys. J. C* **77**, 276 (2017)
- [9] S. L. Glashow, “Partial-symmetries of weak interactions,” *Nucl. Phys.* **22**, 579-588 (1961)
- [10] P. W. Higgs, “Broken symmetries and the masses of gauge bosons,” *Phys. Rev. Lett.* **13**, 508-509 (1964)
- [11] K. Li, “Analysis on the relationship between the Higgs boson and the standard model,” *Theoretical and Natural Science* **38**, 20 (2024)
- [12] Particle Data Group, “Review of Particle Physics,” *Prog. Theor. Exp. Phys.* **2022**, 083C01 (2022)
- [13] A. Djouadi, “The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model,” *Phys. Rept.* **459**, 1 (2008)
- [14] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider,” *JINST* **3**, S08003 (2008)

- [15] ATLAS Collaboration, “Software & Computing - Intel Case Study,” ATLAS Experiment at CERN website. (2023)
- [16] ATLAS Collaboration, “ATLAS b-jet identification performance and efficiency measurement with  $t\bar{t}$  events in pp collisions at  $\sqrt{s} = 13$  TeV,” Eur. Phys. J. C **79**, 970 (2019)
- [17] ATLAS Collaboration, “Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC,” Eur. Phys. J. C **79**, 375 (2019)
- [18] ATLAS Collaboration, “Jet reconstruction and performance using particle flow with the ATLAS Detector,” Eur. Phys. J. C **77**, 466 (2017)
- [19] ATLAS Collaboration, “Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector,” Phys. Lett. B **784**, 173-191 (2018)
- [20] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, “Jet substructure as a new Higgs search channel at the LHC,” Phys. Rev. Lett. **100**, 242001 (2008)
- [21] CMS Collaboration, “CMS Event Display CMS-PHO-EVENTS-2022-018,” CERN Document Server (2022).
- [22] M. Cacciari, G. P. Salam and G. Soyez, “The anti-kt jet clustering algorithm,” JHEP **04**, 063 (2008)
- [23] ATLAS Collaboration, “Search for charged Higgs bosons decaying into a top and a bottom quark in multilepton final states in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector,” ATLAS-CONF-2024-002 (2024)
- [24] E. G. Reynolds, K. F. DiPetrillo and C. Vernieri, “Searches for New High-Mass Resonances Decaying to Fermions at the LHC,” Annu. Rev. Nucl. Part. Sci. **74**, 375 (2024)
- [25] ATLAS Collaboration, “Measurements of the production cross section of a  $Z$  boson in association with jets in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector,” Eur. Phys. J. C **77**, 361 (2017)
- [26] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector,” CERN Yellow Rep. Monogr. **2**, 1-275 (2017)
- [27] ATLAS Collaboration, “Performance of the ATLAS missing transverse momentum reconstruction in proton-proton collisions at  $\sqrt{s} = 13$  TeV,” Eur. Phys. J. C **78**, 903 (2018)
- [28] Particle Data Group, “Review of Particle Physics,” Prog. Theor. Exp. Phys. **2024**, 083C01 (2024)
- [29] A. Radovic et al., “Machine learning at the energy and intensity frontiers of particle physics,” Nature **560**, 41 (2018)
- [30] D. Guest, K. Cranmer and D. Whiteson, “Deep Learning and its Application to LHC Physics,” Ann. Rev. Nucl. Part. Sci. **68**, 161 (2018)

- [31] P. Baldi, P. Sadowski and D. Whiteson, “Searching for exotic particles in high-energy physics with deep learning,” *Nature Communications* **5**, 4308 (2014)
- [32] G. Cerri, S. Duarte, T. Q. Nguyen, M. Pierini, M. Spiropulu and J. R. Vlimant, “Variational Autoencoders for New Physics Mining at the Large Hadron Collider,” *JHEP* **05**, 036 (2019)
- [33] Z. Zhang et al., “Physics-informed deep learning method for predicting tunnelling-induced ground deformations,” *Acta Geotechnica* **18**, 4957 (2023)
- [34] I. Goodfellow, Y. Bengio and A. Courville, “Deep Learning,” MIT Press (2016)
- [35] D. E. Rumelhart, G. E. Hinton and R. J. Williams, “Learning representations by back-propagating errors,” *Nature* **323**, 533-536 (1986)
- [36] A. Paszke et al., “Automatic differentiation in PyTorch,” 31st Conference on Neural Information Processing Systems (2017)
- [37] D. López Robles, “Diagnóstico de miopía patológica en imágenes de fondo de ojo mediante aprendizaje profundo,” Final Degree Project, Universitat Oberta de Catalunya (2020).
- [38] L. Bottou, “Large-Scale Machine Learning with Stochastic Gradient Descent,” Proceedings of COMPSTAT’2010, 177-186 (2010)
- [39] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” arXiv:1412.6980 (2014)
- [40] N. Srivastava et al., “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *Journal of Machine Learning Research* **15**, 1929-1958 (2014)
- [41] S. Rohrmanstorfer et al., “Image Classification for the Automatic Feature Extraction in Human Worn Fashion Data,” *Mathematics* **9**, 624 (2021)
- [42] A. Vaswani et al., “Attention Is All You Need,” *Advances in Neural Information Processing Systems* **30**, 5998-6008 (2017)
- [43] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE* **86**, 2278-2324 (1998)
- [44] J. Hu, L. Shen and G. Sun, “Squeeze-and-Excitation Networks,” *IEEE Conference on Computer Vision and Pattern Recognition*, 7132-7141 (2018)
- [45] M. Lin, Q. Chen and S. Yan, “Network in Network,” arXiv:1312.4400 (2013)
- [46] M. Feickert and B. Nachman, “A guide to applying machine learning in particle physics,” *Rev. Phys.* **7**, 100067 (2021)