Enhancing New Physics Discovery Through Machine Learning: A Case Study for Particle Physics Using the Super Symmetry (SUSY) Dataset

Professional Masters in Data Science and Leadership Programme - 2023

Acelya Deniz Gungordu

May 2025

# Contents

1. [**Chapter 1: Introduction**](#_uwm5t47ah9ue) **2**
   1. [Reflective Account for Chapter 1](#_51djsuergyht) . . . . . . . . . . . . . . . . . . . . . . 3
2. [**Chapter 2: Terms of Reference and Literature Review**](#_djen5pvm6bi) **5**
   1. [2.1 Introduction and Context](#_btx6c2yqbgab) . . . . . . . . . . . . . . . . . . . . . . . . 5
   2. [2.2 Project Aims and Objectives](#_wu7iqd3wdd36) . . . . . . . . . . . . . . . . . . . . . . . 6
      1. [2.2.1 Aims](#_ns00yxg4amww) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
      2. [2.2.2 Objectives](#_84qlspz4i0ba) . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
   3. [2.3 Literature Review](#_saybq8pe6l3i) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
      1. [2.3.1 The Evolution of Analytical Paradigms in Particle Physics](#_j19asn8xxgy0) . 7
      2. [2.3.2 Deep Learning: A New Frontier in High-Energy Physics Data](#_1arf9n6cl786) [Analysis](#_1arf9n6cl786) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
      3. [2.3.3 Addressing the “Null Results” Era: The Rise of Model-Agnostic](#_drvg213twu66)

[and Anomaly Detection Techniques](#_drvg213twu66) . . . . . . . . . . . . . . . . . 8

* + 1. [2.3.4 Enduring Challenges, Best Practices, and the Integrating Role](#_7sv4dhetoj9j)

[of Action Research](#_7sv4dhetoj9j) . . . . . . . . . . . . . . . . . . . . . . . . . . 9

* + 1. [2.3.5 Positioning the Present Research within the Evolving Landscape](#_dngfqugmwly0) 9
  1. [2.4 Research Statement and Research Questions](#_rqunrk5gocy5) 10
     1. [2.4.1 Research Statement](#_6ysl6x4sgmag) 10
     2. [2.4.2 Action Research Question](#_wynbm1a40e00) 11
     3. [2.4.3 Additional Research Questions](#_ecyyrmabr392) 11
  2. [Chapter 2 Reflective Account](#_xomzqjxglcib) 13

1. [**Chapter 3: Methodology**](#_cqph15puxho3) **14**
   1. [Introduction](#_yr65lsy6uigs) 14
   2. [Research Design and Approach](#_hlgh3taskfuj) 14
   3. [Data Collection and Acquisition](#_s79zsix3jdr5) 15
      1. [Justification for Data Source](#_sl9sfqayqrd8) 15
      2. [Dataset Characteristics (SUSY Dataset)](#_fnw49dugtfv2) 16
   4. [Ethical Considerations](#_dza7gz3z4e8n) 16
   5. [Data Science and Machine Learning Methodology](#_rhwgdl8wuadv) 17
      1. [Exploratory Data Analysis (EDA) and Visualization](#_bmno091vyma6) 17
      2. [Data Preprocessing and Transformation](#_7rcqjr4pr7es) 18
      3. [Machine LearningModel Development and Training](#_76suw6kw8xli) 19
      4. [Model Evaluation and Performance Analysis](#_bcohm9acaqoa) 19
      5. [Best Model Evaluation, Hyperparameter Tuning, and K-Fold Cross-](#_4mymqgtzulh6) [Validation](#_4mymqgtzulh6) 20
   6. [Triangulation of Methods](#_w0mio2ptcsrv) 20
   7. [Chapter 3: Reflective Account](#_2zu0p8a7ndfh) 21
2. [**Chapter 4: Project Activities in Cycles**](#_vsox9i85eoh) **22**
   1. [4.1 Cycle1: Defining and Planning for Strategic Change](#_s2529fz0z9lq) 22
   2. [4.2 Cycle2: Refinement, Scaling, and Organisational Impact](#_5lv0cvvjccly) 23
   3. [4.3 Cumulative Evaluation and Personal Impact](#_lr7b0prhrcew) 23
   4. [Chapter 4: Reflective Account](#_u0nmz0jhrlek) 24
3. [**Chapter 5: Project Findings and Analysis**](#_4bmcg4azulea) **25**
   1. [Exploratory Data Analysis (EDA): A Foundational Dialogue with the Data](#_5ldboadg44wz) 25
      1. [Dataset Structure and Class Balance: A Fortunate Starting Point](#_a1d3axxr2ukg) 25
      2. [Feature Distributions: Unveiling Diversity and the Necessity for](#_kdwax9s47uoj) [Scaling](#_kdwax9s47uoj) 26
      3. [Correlation Analysis: Mapping Inter-Feature Relationships](#_iwkfhrzbqze4) 27
   2. [Establishing Baselines: Logistic Regression and XGBoost](#_c38elqr51v36) 29
      1. [Logistic Regression](#_is3vsidvynrh) 29
      2. [XGBoost](#_3riiw4nx4rb9) 30
   3. [Advancing to Deep Learning: The Core Investigation](#_tm4g1qfukhcz) 32
      1. [DNN Architecture and Rationale](#_aasoxk2zuqag) 32
      2. [DNN Performance and Visual Analysis](#_n13fu9xlau64) 32
   4. [Final Synthesis: Comprehensive Model Comparison](#_5w1f2doadrf4) 33
   5. [Chapter 5 Reflective Account](#_aymjgojcedp7) 34
4. [**Chapter 6: Conclusions and Recommendations**](#_y784jnjgtwep) **35**
   1. [6.2 Recommendations](#_fic01vw6xxqy) 37
   2. [6.3 Final Reflection on the Action Research Journey](#_8yliuxe9oqdk) 38
   3. [Chapter 6 Reflective Account](#_6en0g04eiq18) 40

**Project Summary**

The search for new fundamental particles, such as those predicted by Supersymmetry (SUSY), is a primary goal of modern high-energy physics, relying on identifying faint “signal” events within colossal volumes of “background” data. This project aimed to systematically investigate and optimize advanced machine learning techniques, particularly Deep Learning, to enhance the detection of these elusive signals. Using the public benchmark SUSY dataset, the investigation sought to establish a robust, reproducible framework for model comparison while exploring the synergy between automated feature learning and traditional, physicist-engineered features, all within an Action Research framework emphasizing iterative development and structured reflection. To achieve this, the project followed a rigorous quantitative data science pipeline, beginning with an Exploratory Data Analysis of the 5-million-instance dataset. Performance benchmarks were then established using a regularized Logistic Regression and an XGBoost model before a custom Deep Neural Network (DNN) was developed and optimized. All models were rigorously evaluated on a held-out test set using the Area Under the ROC Curve (ROC AUC) for direct comparison. The empirical investigation yielded several key conclusions, chief among them the clear superiority of the custom-designed DNN, which achieved a top-performing ROC AUC of 0.8779. A clear performance hierarchy was established, with the DNN outperforming Logistic Regression, which, surprisingly, outperformed the tuned XGBoost model. For machine learning researchers, future efforts should focus on closing the costs “sim-to-real” gap by applying these methods to real experimental data and exploring more specialized architectures like Graph Neural Networks.

# Chapter 1: Introduction

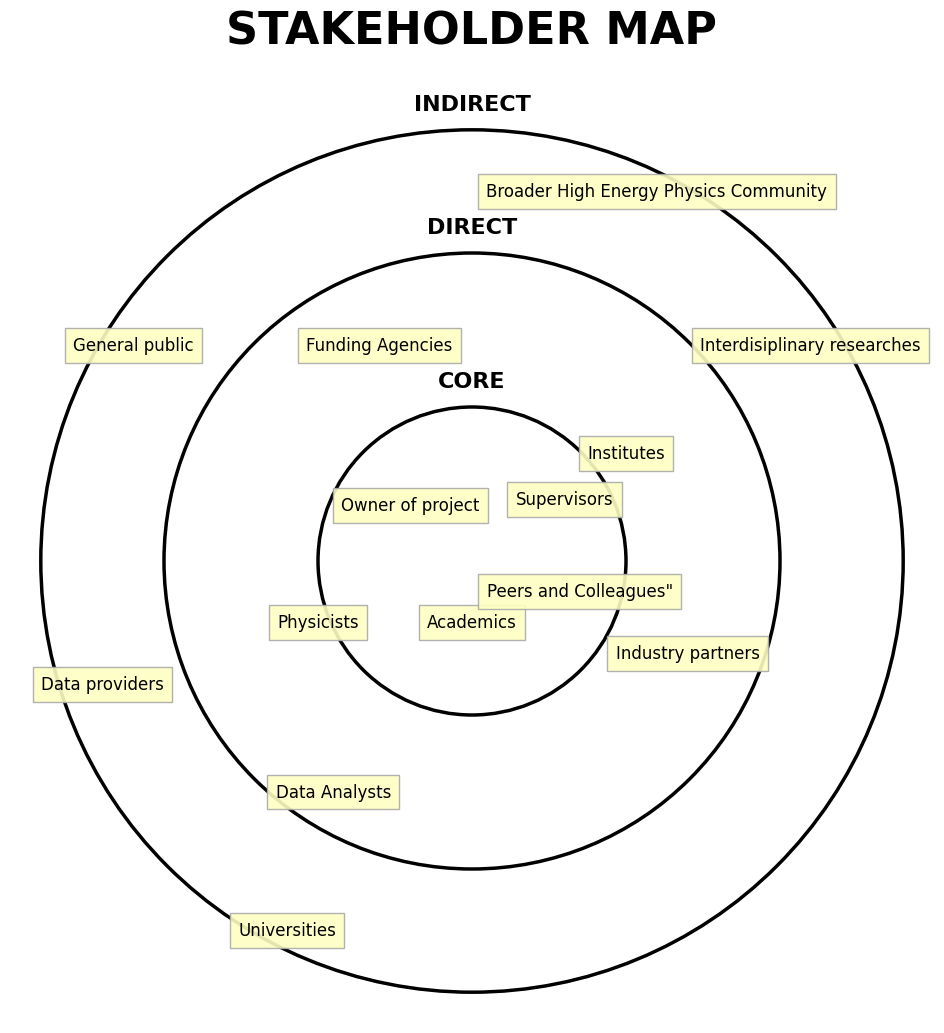
Ever since I first gazed up at the night sky, I have been fascinated by the mysteries of the universe. That youthful curiosity has led me into the worlds of physics and data science, and ultimately to a personal passion: searching for signs of new fundamental particles hidden within massive experimental datasets. For my final project, I decided to challenge myself by working on one of the most active and complex frontiers—using machine learning to identify elusive supersymmetric signals in high energy physics data. My main source of inspiration came from two influential papers: [Mehta et al.](#_dq69pmruaco2) ([2019](#_dq69pmruaco2))

*A high-bias, low-variance introduction to machine learning for physicists* and [Baldi et](#_2mmide2mi6z7) [al.](#_2mmide2mi6z7) ([2014](#_2mmide2mi6z7)) *Searching for exotic particles in high-energy physics with deep learning*. [Mehta](#_dq69pmruaco2) [et al.](#_dq69pmruaco2) ([2019](#_dq69pmruaco2)) offer a uniquely physicist-centered introduction to both modern and foundational machine learning concepts, as well as concrete tutorials applying methods to real physics datasets, including the SUSY benchmark. [Baldi et al.](#_2mmide2mi6z7) ([2014](#_2mmide2mi6z7)) show—using the same SUSY dataset—that deep neural networks dramatically outperform classical “shallow” machine learning models and even engineered features in distinguishing rare signals from background processes in simulated LHC data.

Inspired by their work, my project set out to go beyond published baselines: to **systematically compare and improve classical machine learning techniques and deep learning architectures—including experimenting with different hyperparameters and structures—using the SUSY dataset as a testbed**. My focus was not just on reproducing published results, but on leading a sustained, reflective inquiry into the process itself: learning how and why certain architectures perform better, under what data conditions, and what lessons for real-world discovery in high-energy physics might be drawn ([Baldi et al.](#_2mmide2mi6z7), [2014](#_2mmide2mi6z7); [Mehta et al.](#_dq69pmruaco2), [2019](#_dq69pmruaco2)).

This is not only an academic exercise for me—I see it as a process of action research and data science leadership. Action research, as pioneered by [Lewin](#_77l0cxnjrjmg) ([1946](#_77l0cxnjrjmg)) and developed by [Reason and Bradbury](#_zca1lodqv3tr) ([2001](#_zca1lodqv3tr)), emphasizes a cyclical approach of planning, acting, observing, and reflecting—aiming both for practical improvements (in my case, in the performance and interpretability of SUSY classifiers) and personal growth. Throughout this project, I have tried to maintain a daily reflective diary, documenting not only technical steps but also my emotions, doubts, and evolving leadership mindset. By making my process transparent and iterative, I am enacting the ideals of action learning: confronting ambiguity, surfacing my values and assumptions, and seeking genuine learning for myself and, hopefully, my peers ([Coghlan and Brannick](#_boycjbtwcpkd), [2014](#_boycjbtwcpkd)).

Stakeholder engagement is integral to my approach: while the direct stakeholders for this technical study may be data scientists, physicists, and the broader high energy physics community, I see value in connecting to a wider interdisciplinary audience. Im- proved methods for parsing big, noisy data have value well beyond particle physics—in medicine, industry, and beyond—thus linking my personal project to much larger societal opportunities and risks. The Stakeholder map can be seen in schema 1.



Schema 1: Stakeholder map.

My central research question is:

*How can different machine learning techniques—including modern deep learning architectures with various hyperparameters—be optimized and compared for signal-vs-background classification in the SUSY dataset, and how do these improvements enhance the discovery potential for supersymmetric particles in high energy physics experiments?*

By building on, reproducing, and advancing the approaches outlined in [Mehta et al.](#_dq69pmruaco2) ([2019](#_dq69pmruaco2)) and [Baldi et al.](#_2mmide2mi6z7) ([2014](#_2mmide2mi6z7)), I hope to contribute not only incremental improvements

in classification accuracy, but also deeper insight into the leadership challenges that face data-driven discovery in modern science. Throughout, I am reflecting on my role not just as an analyst, but as an action researcher and leader in a rapidly evolving scientific context ([Northouse](#_7zj6c5udui9r), [2021](#_7zj6c5udui9r)).

Answering this question involves integrating insights from both physics and data science, highlighting the value of interdisciplinary collaboration. The project’s methods are well aligned with contemporary trends that enlist machine learning for tackling high- volume, intricate data. By focusing on a particularly challenging scenario—distinguishing elusive supersymmetric signals within HEP datasets—this endeavor exemplifies the potential and necessity of advanced analytical techniques in modern research ([LeCun et al.](#_3ed3tp3wh8bp), [2015](#_3ed3tp3wh8bp)).

In the chapters that follow, a brief overview of supersymmetry will be provided. Subsequently, the discussion will shift to the methodological aspects of deep learning, outlining crucial model-building and evaluation strategies specific to HEP data. Finally, potential implications of this work for the scientific community will be considered, particularly how successful detection methods might shape future experimental designs, inspire refinements in data analysis, and possibly guide new theoretical perspectives. Ultimately, this project seeks to strengthen the synergy between particle physics and advanced computational approaches, contributing to a deeper comprehension of the universe and illustrating the powerful role of machine learning in driving scientific breakthroughs.

## Reflective Account for Chapter 1

As I navigate my journey from physics into data science, I find myself reflecting on what I am intentionally including and excluding in my approach, not just for my own learning, but also in relation to the “second and third persons” – my peers, the scientific community, and future practitioners in this interdisciplinary space. My domain interest is deeply rooted in the methods and mindset of physics: I am inherently quantitative and drawn to big, fundamental questions. However, as I adapt to the data science field, I am learning to balance scientific curiosity with the practical considerations and broader impacts that are integral to action research.

In this project, I am intentionally including the application of rigorous, evidence-based machine learning techniques to a physics-motivated problem: the search for supersymmetric particles using benchmark open datasets such as SUSY. I bring with me a physicist’s analytical skills, as well as commitment to grounding decisions in the existing scientific literature. As an action researcher, however, I am also including a process of transparent reflection, documenting my decisions, challenges, and learning in a personal diary. This self-observation is not solely for my benefit; it is part of the action research commitment to generate actionable knowledge that might be useful to others facing similar interdisciplinary challenges ([Coghlan and Brannick](#_boycjbtwcpkd), [2014](#_boycjbtwcpkd); [Reason and Bradbury](#_zca1lodqv3tr), [2001](#_zca1lodqv3tr)). By sharing my approach and lessons learned, I hope to contribute to a collective understanding within our learning community and potentially the broader data science in physics community.

Yet, I have also had to determine what to exclude. Out of necessity, I am not generating new raw data or seeking direct access to experimental collaborations; instead, I work with openly available, pre-processed datasets. While this limits the project’s scope and my direct engagement with the data generation and data taking processes, it makes the work more accessible and potentially reproducible for others using the same public

resources. My project does not encompass the design of entirely novel deep learning algorithms but rather focuses on the application, optimization, and comparison of established tools in a well-defined physics context. This focus on applying and evaluating existing methods is perhaps a more realistic initial contribution that can inform the practice of others in the field.

Through honest self-examination, I am becoming aware of possible biases and blind spots in my approach. My own limited hands-on experience with complex deep learning models is a primary blind spot. While I understand the theoretical concepts, the practical skill and intuition needed for effective hyperparameter tuning and model selection are still developing. This lack of practical intuition is a common challenge for students entering the field, highlighting a potential area where shared learning and transparent documentation (part of the action research cycle) can benefit others.

Furthermore, I quickly became aware of the inherent difficulty and depth of this specific topic – particle physics signal detection is a field with decades of specialized expertise. It is easy for someone new to the field to overlook subtle but crucial aspects that seasoned experts understand intuitively. This likely contributed to an initial underestimation of the project timeline and the complexities involved. These are not just my personal struggles; they reflect systemic challenges in interdisciplinary work and the difficulty for newcomers (like myself) to quickly grasp implicit knowledge. My experience, documented reflectively, can perhaps serve as a cautionary tale or a guide for others navigating similar learning curves.

At each stage, I am discovering new perspectives and ways of thinking that resonate beyond my individual experience. I have come to appreciate that defining a unique research question and path as a student is itself a significant challenge, requiring independent thought and decision-making under uncertainty. This process, while personal, reflects a broader need for adaptability and initiative in the rapidly evolving landscape of data science – skills vital for future leaders in any field ([Northouse](#_7zj6c5udui9r), [2021](#_7zj6c5udui9r)).

I am also discovering the iterative nature of action research firsthand. It’s not a linear process, but rather a continuous cycle of trying something (action), seeing the result (observation), and then stepping back to understand what happened and what to do next (reflection and planning). This is teaching me patience and adaptability, pushing me beyond just seeking the ”right answer” to understanding the messy reality of research and development. This cyclical learning process, documented in my diary, generates a personal theory of action ([Argyris and Schon](#_vajso9ep3j0z), [1974](#_vajso9ep3j0z)) that I can share, offering concrete examples of how reflection fuels progress, not just for me, but potentially for others in similar learning situations.

Through this process, several shortcomings in my practice and the broader system have become apparent. One major challenge is the difficulty of obtaining real, representative data from physics collaborations or experiments for external student research. This logistical hurdle necessitated pivoting to using publicly available benchmark datasets. This experience highlighted a significant real-world challenge in data science: data access and wrangling are often the hardest parts, and the reliance on easily available data can constrain the kinds of problems individuals or small teams can address. This isn’t just my personal drawback; it points to a systemic issue regarding data accessibility that impacts many researchers.

Furthermore, as I begin to form my own opinions and interpretations based on my analysis, I am becoming aware of the potential for bias against student perspectives. Gaining acknowledgment or validation for nascent insights from established experts can be

challenging due to perceived inexperience or lack of domain authority. This underscores the importance of rigorous methodology, clear communication, and humility in presenting findings. It also highlights a potential barrier to innovation if established fields are too resistant to perspectives from those applying methods from other domains. My effort to conduct rigorous work and articulate its value is, in a small way, an action aimed at navigating and perhaps slightly pushing back against this systemic bias.

Altogether, this project is proving to be a rich learning experience, not only technically, but also personally and professionally. It is challenging my assumptions, revealing my blind spots, and forcing me to grow as a researcher and a burgeoning data science leader navigating complexity and uncertainty. More than just a personal endeavor, it is an exercise in action research, highlighting how individual learning and challenges con- nect to broader issues of knowledge generation, practice, and systemic realities within interdisciplinary fields.

# Chapter 2: Terms of Reference and Literature Re- view

This chapter establishes the foundational groundwork for this action research project. It commences by outlining the context of high-energy physics research and the specific challenges in detecting Supersymmetry, thereby setting the stage for the project’s aims and objectives. A comprehensive literature review follows, charting the evolution of analytical techniques in the field, highlighting the rise of deep learning, and underscoring the rationale for adopting an action research methodology. Subsequently, the chapter articulates the precise research statement and guiding questions, defines the project’s boundaries and scope, and concludes with an initial reflective account, integral to the action research process, capturing the researcher’s perspective at this juncture.

## 2.1 Introduction and Context

High-energy physics (HEP) experiments, particularly those conducted at facilities like the Large Hadron Collider (LHC), are monumental endeavors designed to probe the fundamental structure of matter and the forces that govern it ([CERN](#_rtggqxk5jny5), [2020](#_rtggqxk5jny5)). These experiments generate colossal volumes of collision data, within which physicists search for rare signals indicative of new physical phenomena. Among the numerous theoretical frameworks proposed to extend the Standard Model (SM) of particle physics, Supersymmetry (SUSY) remains a particularly compelling candidate ([Wess and Zumino](#_f875x9va4naz), [1974](#_f875x9va4naz); [Kane](#_tro9c8kpca3), [2017](#_tro9c8kpca3)). SUSY postulates a fundamental symmetry between fermions and bosons, predicting a superpartner for every known SM particle. The discovery of these super- symmetric particles would not only validate this elegant theory but also potentially offer solutions to long-standing puzzles such as the nature of dark matter and the hierarchy problem ([Ellis](#_p2mwomv49ljd), [2016](#_p2mwomv49ljd)).

However, the search for SUSY is fraught with challenges. If these particles exist at energies accessible to current colliders, their production rates are likely to be extremely low, and their decay signatures are often subtle and easily lost within the overwhelming back- ground of common SM processes. Traditional analysis methods, while foundational, often struggle with the sheer complexity and high dimensionality of HEP data, necessitating the development and adoption of novel approaches that can effectively discern these faint

signals from the noise. Recent advances in machine learning, especially deep learning, have shown immense promise in automatically learning complex, non-linear patterns directly from raw data, offering new avenues for discovery. This project specifically focuses on leveraging these advanced methodologies, using the publicly available SUSY dataset from the UCI Machine Learning Repository as a benchmark testbed, to enhance the detection performance for supersymmetric signals. This endeavor is framed within an *action research methodology*, emphasizing an iterative cycle of planning, acting, observing, and reflecting to achieve both practical improvements in model performance and deeper research insights.

## 2.2 Project Aims and Objectives

The overarching purpose of this project is to explore and enhance the application of machine learning and advanced deep learning models for the classification of signal versus background processes within the context of the SUSY dataset, thereby contributing to the broader effort of improving the detection capabilities for supersymmetric particles in HEP.

### 2.2.1 Aims

The primary aims of this project are:

* + - * To systematically investigate the efficacy of advanced deep learning models in enhancing the classification accuracy of signal (SUSY) versus background processes using the benchmark SUSY dataset.
      * To demonstrate and critically evaluate the application of an action research approach within a data science context, illustrating how iterative refinement and reflective practice can lead to improved research insights and practical model performance in a complex scientific domain.

### 2.2.2 Objectives

To realize these aims, the project is structured around specific objectives targeted at different beneficiaries, reflecting the multifaceted nature of action research:

### For Me (the Researcher):

* + - * To develop a robust and reproducible experimental framework for designing, training, and evaluating deep neural networks on high-energy physics data, specifically addressing challenges such as imbalanced classes and high-dimensional feature spaces.
      * To systematically optimize model architectures and hyperparameters to achieve maximum classification accuracy, effectively leveraging physics-inspired features.
      * To cultivate a deeper understanding of applying deep learning techniques within HEP research, thereby refining skills in data manipulation, model building, rigorous evaluation, and the interpretation of results within this complex scientific field.
      * To enhance project management, problem-solving, and interdisciplinary communication skills through the iterative cycles inherent in action research.

### For Us (the Project Team/Research Community):

* + - * To create and disseminate a reproducible research prototype (code, models, and findings) that can serve as a valuable baseline and stimulus for future explorations of advanced deep learning methods in high-energy physics.
      * To contribute to the ongoing research in particle physics by providing a robust and effective framework for identifying supersymmetric signals in collision data, potentially informing strategies for future experimental analyses and aiding in the discovery of new particles.
      * To generate and share insights into effective strategies for automatic feature extraction and model optimization in complex, high-dimensional data domains, benefiting the broader machine learning and data science communities engaged in scientific discovery.

### For Others (the Wider Scientific and Industrial Community):

* + - * To demonstrate the broader applicability and power of deep learning techniques in diverse scientific domains, particularly those characterized by large, complex datasets with subtle signal patterns and significant noise.
      * To contribute to the body of literature evidencing the benefits and challenges of deep learning in high-energy physics, offering insights and inspiration for applications in other scientific and industrial fields dealing with complex, high-volume data.
      * To suggest potential improvements or considerations for experimental design and data processing pipelines in future collider experiments, informed by the findings from deep learning model optimization and performance analysis.

## 2.3 Literature Review

### 2.3.1 The Evolution of Analytical Paradigms in Particle Physics

Historically, HEP analyses relied on physicists’ intuition to define kinematic selections (“cuts”) to isolate regions of interest in the data. While foundational, these “cut-and- count” methods often lack the sensitivity required for complex, high-dimensional signals. The early 2000s witnessed the integration of “shallow” machine learning (ML) algorithms, such as Boosted Decision Trees (BDTs) and Support Vector Machines (SVMs). These methods, typically applied to a set of features meticulously engineered by physicists from theoretical principles and detector knowledge (e.g., invariant masses, event shape variables), offered a significant improvement in discrimination power ([Bhattacherjee and](#_5vp7u28hps3g) [Mukherjee](#_5vp7u28hps3g), [2024](#_5vp7u28hps3g)). The discovery of the Higgs boson, for instance, heavily benefited from the application of BDTs. Despite their successes, these shallow learners are intrinsically limited by the expressiveness of the handcrafted features; crucial information might be sub-optimally represented or inadvertently discarded during the feature engineering process ([Baldi et al.](#_2mmide2mi6z7), [2014](#_2mmide2mi6z7)).

### 2.3.2 Deep Learning: A New Frontier in High-Energy Physics Data Analysis

The remarkable achievements of deep learning (DL) in domains such as computer vision and natural language processing ([LeCun et al.](#_3ed3tp3wh8bp), [2015](#_3ed3tp3wh8bp)) spurred its adoption in HEP. Deep Neural Networks (DNNs) presented a transformative capability: the ability to learn complex, hierarchical representations directly from “raw” data, thereby automating the feature engineering process. A seminal contribution by [Baldi et al.](#_2mmide2mi6z7) ([2014](#_2mmide2mi6z7)), utilizing the same benchmark SUSY dataset that forms the core of the present study, compellingly demonstrated this potential. Their work showed that a relatively simple DNN, trained directly on basic kinematic features, could significantly outperform traditional BDTs operating on sophisticated, physicist-engineered features in distinguishing SUSY signals from background. This was a pivotal moment, signaling that DL could uncover intricate data patterns that might elude manual feature construction.

This breakthrough ignited a flurry of research into various DL architectures tailored for HEP data. Convolutional Neural Networks (CNNs) have been successfully applied to “jet images” and calorimeter data, treating detector outputs as images for tasks like particle identification and jet flavor tagging ([Karagiorgi et al.](#_tro9c8kpca3), [2022](#_tro9c8kpca3)). Recurrent Neural Networks (RNNs) and, more recently, Transformers, are employed for sequential data, such as particle tracks or jet constituents. Graph Neural Networks (GNNs) have emerged as a powerful paradigm for leveraging the inherent relational structure of particle inter- actions and detector geometries, representing particles or detector elements as nodes and their relationships as edges ([Bhattacherjee and Mukherjee](#_5vp7u28hps3g), [2024](#_5vp7u28hps3g); [Suresh et al.](#_k9tsx45jfygb), [2024](#_k9tsx45jfygb)). These advanced architectures have consistently demonstrated enhanced performance in classification accuracy, signal efficiency, and background rejection across a wide spectrum of physics analyses.

### 2.3.3 Addressing the “Null Results” Era: The Rise of Model-Agnostic and Anomaly Detection Techniques

The continued absence of definitive signals for new physics in many targeted searches, despite the increased sensitivity afforded by ML, has prompted a strategic shift towards more model-agnostic and unsupervised approaches. The concern is that new physics might manifest in unexpected regions of phase space or through unforeseen signatures not covered by searches optimized for specific theoretical models like SUSY. This has fueled the development of anomaly detection techniques. Methods such as Classification Without Labels (CWoLa), introduced by [Collins et al.](#_1ld3nj762tx2) ([2018](#_1ld3nj762tx2)), leverage classifiers trained on data from distinct regions (e.g., a signal-rich region versus background-dominated side- bands) to pinpoint discrepancies without explicit signal labels. Density-based approaches, exemplified by the work of [D’Agnolo and Wulzer](#_aaqcjowpli2d) ([2019](#_aaqcjowpli2d)), aim to model the probability density of background events, thereby flagging events in low-density regions as potentially anomalous. Autoencoders and normalizing flows, trained to accurately reconstruct or model SM background events, can identify anomalies as events that are poorly re- constructed or exhibit low likelihood under the learned background model ([Belis et al.](#_awjd98koooze), [2024](#_awjd98koooze); [Suresh et al.](#_k9tsx45jfygb), [2024](#_k9tsx45jfygb)). These methods represent a vital move towards data-driven discovery, lessening the dependence on specific theoretical predictions.

### Enduring Challenges, Best Practices, and the Integrating Role of Action Research

Despite its transformative impact, the application of ML, particularly DL, in HEP is not without significant challenges. **Interpretability** remains a major concern; the ‘black box” nature of many DNNs makes it difficult to understand the physical basis for their classifications, which is essential for scientific validation and building trust within the physics community ([Suresh et al.](#_k9tsx45jfygb), [2024](#_k9tsx45jfygb)). Ensuring robust **generalization** from simulated training data (which invariably contains imperfections) to real experimental data is another critical issue, necessitating meticulous calibration and comprehensive uncertainty quantification ([Karagiorgi et al.](#_tro9c8kpca3), [2022](#_tro9c8kpca3)). The rigorous estimation of **systematic uncertainties** associated with ML-based analyses and establishing the statistical significance of potential excesses identified by these algorithms (often termed the ‘look-elsewhere effect” in high-dimensional search spaces, see [Gross and Vitells](#_9jxx0vukaoav) [2010](#_9jxx0vukaoav)) are ongoing areas of intense research ([Belis et al.](#_awjd98koooze), [2024](#_awjd98koooze)).

In response to these challenges, the HEP community has been actively developing and promoting best practices. These include the adoption of rigorous cross-validation techniques, meticulous hyperparameter optimization, the use of publicly available benchmark datasets such as SUSY and HIGGS to facilitate reproducible research and fair comparison of methods ([Mehta et al.](#_dq69pmruaco2), [2019](#_dq69pmruaco2)), and the transparent reporting of a comprehensive suite of performance metrics.

It is precisely within this dynamic, complex, and evolving landscape that an *action research* methodology offers a uniquely valuable and integrating framework. Action re- search, as conceptualized by [Lewin](#_77l0cxnjrjmg) ([1946](#_77l0cxnjrjmg)) and further elaborated by practitioners like [Reason and Bradbury](#_zca1lodqv3tr) ([2001](#_zca1lodqv3tr)) and [Coghlan and Brannick](#_boycjbtwcpkd) ([2014](#_boycjbtwcpkd)), emphasizes a cyclical process of planning, acting (experimenting), observing (evaluating results), and reflecting (learning and adapting). This iterative paradigm is exceptionally well-suited to intricate problem domains where solutions are not immediately evident and where practical experimentation directly informs and refines theoretical understanding and methodological choices. [Schon](#_rtlxqu5e9k7z) ([1983](#_rtlxqu5e9k7z))’s notion of the “reflective practitioner” further underscores the critical importance of continuous learning, critical self-assessment, and adaptation in professional practice—a mindset indispensable for navigating the rapidly advancing intersection of HEP and ML. In the context of this specific project, action research provides a structured approach not only to develop and optimize sophisticated DL models but also to systematically document and reflect upon the methodological decisions, the learning journey itself, and the broader implications for both the researcher’s technical skill development and their capacity for scientific leadership. It encourages a profound engagement with the problem, fostering an understanding that transcends mere algorithmic application to encompass the human, contextual, and epistemic factors that shape scientific discovery and innovation.

### Positioning the Present Research within the Evolving Landscape

This project is strategically positioned at the confluence of these critical developments in HEP data analysis. It directly builds upon the foundational work of [Baldi et al.](#_2mmide2mi6z7) ([2014](#_2mmide2mi6z7)) by revisiting the well-established SUSY benchmark dataset. The primary technical objective is to leverage contemporary deep learning techniques and a systematic, iterative optimization strategy to rigorously investigate the extent to which these advanced models can enhance signal-versus-background classification accuracy. A particular focus is placed

on evaluating their capacity to automatically extract maximally discriminative features from low-level detector measurements, thereby potentially reducing reliance on, or even surpassing the performance of, traditional handcrafted high-level features, a theme echoed in recent reviews ([Mehta et al.](#_dq69pmruaco2), [2019](#_dq69pmruaco2); [Karagiorgi et al.](#_tro9c8kpca3), [2022](#_tro9c8kpca3)).

However, the ambition of this project extends beyond a mere performance bench- mark. Through the explicit adoption of an *action research methodology* ([Coghlan and](#_boycjbtwcpkd) [Brannick](#_boycjbtwcpkd), [2014](#_boycjbtwcpkd); [Reason and Bradbury](#_zca1lodqv3tr), [2001](#_zca1lodqv3tr)), it aims to meticulously document, analyze, and reflect upon the iterative process of model development, optimization, and evaluation. This encompasses a critical examination of the challenges encountered (e.g., navigating computational constraints, the intricacies of hyperparameter tuning, managing the trade-offs between performance and interpretability) and the learning derived from systematically addressing them. The research will explore how different DL architectures (primarily feed-forward networks, with potential consideration for others if data structure and project scope permit) and regularization techniques impact performance on this specific, yet highly representative, HEP classification problem.

In doing so, this project seeks to address several key research gaps and contribute to ongoing discussions:

* + - * It provides a contemporary and rigorous re-evaluation of deep learning’s capabilities on a foundational HEP benchmark dataset, specifically probing the limits and efficacy of automated feature learning from low-level inputs.
      * It uniquely and explicitly integrates an action research framework into a technical data science project within the HEP domain. This offers valuable insights into the practical challenges, learning opportunities, and personal development inherent in such an interdisciplinary undertaking, aiming to enhance not only the technical out- comes but also the researcher’s problem-solving acumen and leadership capacities in a complex scientific environment.
      * It contributes to the evolving dialogue on best practices for applying ML in physics, particularly concerning robust model selection, effective optimization strategies, and the practicalities of working with high-dimensional, noisy, and often imbalanced datasets.

Ultimately, this research endeavors to deliver both robust technical results pertinent to SUSY signal detection and valuable meta-level insights into the process of conducting data-intensive scientific inquiry through an action research lens. It aims to demonstrate how a reflective, iterative, and critically engaged approach can lead to improved model performance, deeper methodological understanding, and a more holistic appreciation of the interplay between advanced computational methods and the pursuit of fundamental physics.

## 2.4 Research Statement and Research Questions

Building upon the context and literature review, the core inquiry of this project is formally defined as follows:

### 2.4.1 Research Statement

This action research project investigates the development, systematic optimization, and critical evaluation of advanced deep learning techniques to enhance the detection

sensitivity for supersymmetric signals within noisy, high-dimensional High-Energy Physics (HEP) data, using the benchmark SUSY dataset. By directly comparing the performance of various deep learning models with established shallow learning algorithms and by exploring automated feature learning from low-level inputs, this study aims to establish an effective framework for deep network application. The project adopts and champions an action research approach, promoting iterative method refinement, continuous learning, and reflective practice on both technical skills and interdisciplinary leadership within the scientific discovery process.

### 2.4.2 Action Research Question

The central action research question guiding my personal and methodological journey through this project is:

*How can I, as a researcher bridging data science and physics, effectively leverage, adapt, and iteratively refine deep learning methodologies to automatically extract and optimize complex discriminative features from HEP data, thereby demonstrably improving classification accuracy for detecting supersymmetric events, while simultaneously devel- oping my research capabilities and contributing actionable insights through a structured action research process?*

### 2.4.3 Additional Research Questions

To provide a more granular focus and guide the investigation, the project will address the following additional research questions, categorized for clarity:

### Model Development and Optimization:

* + - * How do different configurations of feed-forward deep neural network architectures (e.g., variations in depth, width, activation functions) impact the classification performance for SUSY signal versus background events on the benchmark dataset?
      * To what extent can regularization techniques (e.g., dropout, L1/L2 regularization, batch normalization) effectively mitigate overfitting and enhance the generalization capabilities of deep learning models, given the large volume and inherent complexity of the simulated SUSY dataset?
      * What hyperparameter tuning strategies (e.g., grid search, random search, Bayesian optimization) and specific loss functions (e.g., those tailored for imbalanced datasets) prove most effective and efficient for optimizing the performance of deep learning models applied to this HEP classification task?

### Feature Extraction and Representation Learning:

* + - * To what degree can deep learning models effectively identify and learn the most relevant features directly from the raw SUSY data, without explicit reliance on domain-specific physics knowledge encoded in pre-engineered features?

### Methodological Impact and Action Research Insights:

* + - * How does the iterative nature of the action research cycle (plan-act-observe-reflect), with its emphasis on continuous reflection on biases, assumptions, and blind spots, contribute to the ongoing improvement of the deep learning model’s performance and the overall research strategy within this specific context?
      * How do the results obtained from the developed deep learning models, in terms of signal detection accuracy, computational efficiency, and robustness, compare quantitatively and qualitatively to those achieved with traditional machine learning methods commonly employed in high-energy physics?
      * What are the broader implications of applying these deep learning techniques to HEP data for future experimental designs, data processing pipelines, and theoretical frameworks in particle physics, and how might these advancements influence the potential for discovery significance in future collider experiments?
      * How does the performance of the developed deep learning models scale with variations in the size of the training dataset, and what are the pragmatic computational resource requirements for training and deploying these models in a setting analogous to real-world high-energy physics research?
      * What are the potential limitations, inherent biases (e.g., from simulation, model architecture choices), and interpretability challenges associated with using deep learning models for SUSY particle detection, and how can these be proactively ad- dressed or mitigated to ensure the reliability, scientific validity, and trustworthiness of the results?

## Boundaries and Scope of Investigation

To ensure manageability, depth, and focus within the available resources and timeframe, the scope of this project is deliberately defined by the following considerations:

* **Data Source:** The primary and sole data source for this investigation will be the publicly available SUSY dataset from the UCI Machine Learning Repository. This dataset consists of 5,000,000 Monte Carlo simulated collision events, where the first 8 features are high-level importance kinematic properties measured by particle detectors, and the next 10 features are low-level importance variables de- rived by physicists to help discriminate between the two classes: signal (SUSY) and background. While acknowledging the existence of other relevant benchmark datasets (e.g., the HIGGS dataset), this project will maintain a singular focus on the SUSY dataset to allow for in-depth analysis and model optimization specific to its characteristics.
* **Computational Resources:** The training of deep neural networks, especially on large datasets, is computationally intensive. This project will optimize model devel- opment and experimentation within the constraints of available GPU-accelerated computing platforms. Practical compromises regarding model complexity, the ex- tent of hyperparameter searches, or the fraction of the dataset used for certain exploratory phases may be necessary to ensure feasible training times and project completion.
* **Methodological Focus:** The core methodological focus will be on feed-forward deep neural networks, associated optimization techniques relevant for supervised classification tasks. This includes exploring various network architectures, activa- tion functions, regularization methods, and optimizers. The primary investigation will center on refining and understanding the application of well-established deep

learning approaches for this specific classification problem. A comparative analysis against traditional shallow learning algorithms (e.g., Logistic Regression, BDTs) will be conducted as a baseline.

* **Theoretical Integration and Physics Interpretation:** The underlying physics of Supersymmetry and the principles of particle detection are considered essential background knowledge for interpreting the data, guiding feature selection (if com- paring low-level vs. high-level importance features), and evaluating the physical plausibility of results. However, the project’s primary emphasis is on data-driven model improvement and methodological innovation within the machine learning domain rather than on developing new theoretical physics models, proposing modifications to SUSY theory, or fundamentally altering existing physics paradigms. Interpretation of learned features will be qualitative where possible, but formal development of new interpretable ML methods is beyond the scope.

## 2.5 Chapter 2 Reflective Account

Adopting an action research methodology at this stage of the project has demanded ongoing, critical reflection on my process as both a data scientist and interdisciplinary researcher, and has made me increasingly aware of both what I am intentionally including and what I am excluding to maintain a focused, feasible scope. I have chosen to rigorously explore feed-forward deep learning models—prioritizing both low-level and high-level features—since the literature and the structure of the SUSY dataset make these most appropriate for benchmarking automated feature learning, regularization, and systematic hyperparameter tuning. This focus necessarily means I am excluding comparative analysis with alternative BSM theories, as well as more specialized deep learning architectures such as graph neural networks or transformers, and I treat classic ML methods as baselines rather than subjects of deep optimization. While this pragmatic narrowing keeps the work manageable, I recognize it risks reinforcing my own bias towards deep learning solutions—a bias also fed by my tendency, as data scientist, to prioritize model performance metrics over nuanced physical interpretability. These biases, and additional blind spots such as undervaluing the complexities of the simulation process, the interpretability challenges posed by deep models, or the true resource requirements of exhaustive searches, have become clear through both literature review and cycles of trial and error. Such realizations have shaped my current mindset: rather than seeing obstacles as failures, I have begun to internalize that much of scientific progress is non-linear and iterative—where feedback, adaptation and transparent documentation are keys to meaningful improvement. Still, I am increasingly aware of the project’s drawbacks, including the “black box” nature of deep learning models that complicates communication with physicist audiences, as well as time- and resource-intensive model training which sometimes necessitates strategic feature selection or sampling. In addition, limited access to real experimental datasets restricts the external validity and realism of my findings—a systemic challenge for many scientists in my position. Despite these constraints, this project is pushing me to develop genuine leadership and problem-solving capabilities, as I break complex goals into actionable steps, communicate with and seek feedback from diverse audiences, and iterate based on both evidence and self-reflection. I am learn- ing that my main strength lies in bridging disciplines, rigor, and reflective practice, and that my impact is not just in producing a technically robust pipeline, but in modelling

openness and adaptability—making the learning process itself explicit and valuable. Ultimately, seeing my work as small step towards more open, reproducible, and participative data science in high-energy physics both motivates and grounds my research, ensuring that my personal growth is interwoven with the project’s technical and methodological achievements.

# Chapter 3: Methodology

## Introduction

This chapter details the methodological approach employed in this master’s project, out- lining the research design, data collection strategies, and analytical techniques utilized to address the research objectives. The aim is to provide a transparent and justifiable ac- count of the procedures undertaken, allowing for critical evaluation and replication. This section specifically addresses the choice of research paradigm, the practical considerations for data acquisition, and the comprehensive data science pipeline, from Exploratory Data Analysis to advanced Machine Learning model deployment and evaluation.

## Research Design and Approach

The research undertaken in this project primarily adopted a **quantitative, experimental research design**, with a strong emphasis on **predictive modeling** within the domain of data science. This approach was chosen to systematically evaluate the performance of various Machine Learning algorithms on a complex, high-dimensional dataset, with the goal of identifying an optimal model for a specific classification task.

While a traditional qualitative or mixed-methods approach is not directly applicable to the core Machine Learning task, the overall project lifecycle, particularly the iterative nature of model development, refinement, and performance tuning, bears resemblances to aspects of **Action Research**. In this context, Action Research can be understood as a participatory and iterative process of inquiry that seeks to bring about change and improvement in a specific practice or domain through a cyclical process of planning, acting, observing, and reflecting. For a practitioner-researcher in data science, this translates to:

* **Planning:** Defining the problem, selecting data, and outlining initial modeling strategies.
* **Acting:** Implementing data preprocessing, training models, and conducting experiments.
* **Observing:** Analyzing model performance, identifying limitations, and under- standing failure modes.
* **Reflecting:** Interpreting results, refining hypotheses, and deciding on next steps (e.g., trying a different model, feature engineering, hyperparameter tuning).

The suitability of action research for conducting practitioner research in data science lies in its flexible and adaptive nature. **Pros of Action Research in Data Science:**

* **Iterative Improvement:** It naturally accommodates the iterative development cycle inherent in data science, allowing for continuous refinement of models and approaches based on observed outcomes.
* **Practical Relevance:** It emphasizes practical problem-solving and immediate application of findings, aligning well with the goal of deploying effective predictive models.
* **Learning by Doing:** It fosters a deeper understanding of the problem and data through hands-on engagement and continuous feedback loops.
* **Flexibility:** It permits adjustments to the methodology as new insights emerge from the data, which is crucial in often unpredictable data science projects.

### Cons of Action Research in Data Science:

* **Subjectivity/Bias:** The close involvement of the researcher can introduce biases if not carefully managed through rigorous validation techniques.
* **Generalizability:** Findings might be highly specific to the dataset or problem context, potentially limiting their generalizability to other scenarios without further validation.
* **Scope Creep:** The iterative nature can sometimes lead to an expansion of the project scope if not properly constrained.
* **Documentation Overhead:** Requires meticulous documentation of each iteration and decision to maintain rigor and transparency.

In this project, the principles of action research underpinned the iterative refinement of the Machine Learning pipeline, particularly during the model selection and optimization phases (e.g., hyperparameter tuning using GridSearch and cross-validation, which involve systematic experimentation and reflection on results). However, the primary focus remained on systematic quantitative evaluation rather than direct intervention within a real-world system.

This methodological choice is justified by the project’s objective to rigorously compare and contrast the efficacy of different Machine Learning algorithms on a specific dataset. The emphasis is on empirical performance measurement, statistical validation, and reproducible results, which are hallmarks of a quantitative experimental design.

## Data Collection and Acquisition

The data for this master’s project was sourced from the [UCI Machine Learning Repository](https://archive.ics.uci.edu/), specifically the **SUSY dataset** ([Whiteson](#_pv0hvyqszebq), [2014](#_pv0hvyqszebq)). This choice was made due to several practical and strategic considerations, particularly given the constrained time- frame and the independent nature of this research.

### Justification for Data Source

1. **Accessibility and Convenience:** As a single-author project with a limited time- line, acquiring, curating, and validating novel experimental data from a complex domain like particle physics (e.g., from CERN’s CMS experiment) would have been exceptionally time-consuming and required extensive collaborative involvement. The UCI Machine Learning Repository provides readily available, clean, and pre-processed datasets.
2. **Prior Approval and Public Availability:** The SUSY dataset is derived from Monte Carlo simulations related to the CMS experiment, and critically, it is part of the CMS Open Data initiative. This ensures that the data is approved for public use and research outside of the formal CMS collaboration. This eliminates concerns regarding data confidentiality and intellectual property rights, simplifying the ethical and logistical overheads.
3. **Pre-tagging and Quality:** The dataset is well-structured, with features and la- bels (signal vs. background) clearly defined and pre-tagged. This significantly reduced the need for extensive initial data cleaning and feature engineering, allowing the project to focus directly on the core Machine Learning challenges. This convenience was also highlighted by my supervisors, making it a pragmatic choice for demonstrating advanced Machine Learning capabilities.
4. **Relevance to Research Area:** The dataset originates from high-energy physics, aligning with the complex scientific domain that often drives the development of advanced data science techniques. This provides a rich context for demonstrating the application of Machine Learning.

### Dataset Characteristics (SUSY Dataset)

The analysis in this report is conducted on the Supersymmetry (SUSY) dataset, sourced from the UCI Machine Learning Repository. Originally presented by Baldi et al. (2014), the dataset is designed for a binary classification task: distinguishing between signal events, which correspond to a hypothetical supersymmetric process, and background events, which are consistent with the Standard Model of particle physics. The dataset contains 5,000,000 instances, each representing a simulated high-energy collision created using a standard particle physics pipeline (MadGraph, PYTHIA, and DELPHES) to model conditions at the Large Hadron Collider (LHC). Each event is characterized by 18 numerical features. The first eight of these are high-level importance of low-level kinematic properties directly measured from the collision’s final state, such as the momentum (pT), pseudo-rapidity (eta), and angle (phi) of two leptons, as well as the magnitude of missing energy. The remaining ten features are low level importance of high-level variables derived from these initial measurements. These have been specifically engineered by physicists to enhance the discrimination between signal and background events, making them powerful predictors for the classification task. A key advantage of this dataset is that it is complete with no missing values, simplifying the preprocessing pipeline. This controlled and readily available dataset allowed for a focused and robust exploration of various Machine Learning methodologies without being encumbered by the intricacies of raw data acquisition from a live experimental setup.

## Ethical Considerations

Given the use of publicly available and open-source data (CMS Open Data, hosted on UCI Machine Learning Repository) derived from Monte Carlo simulations, the project involved no direct collection of personal, sensitive, or confidential human data. The data is anonymized, aggregated, and intended for public scientific research. Therefore, specific ethical concerns related to data privacy, informed consent, or potential harm to individuals are not applicable.

Table 1: Description of Kinematic Features in the SUSY Dataset (Compact)



### Feature Description



#### High-Level importance Features (Low Level kinematics-directly Measured)

pT: Transverse momentum

Lepton 1

Lepton 2

eta: Pseudo-rapidity

phi: Azimuthal angle

pT: Transverse momentum

eta: Pseudo-rapidity

phi: Azimuthal angle



Missing Energy magnitude: Magnitude of missing transverse momentum

phi: Azimuthal angle of missing transverse momentum



#### Low-Level importance Features (High level kinematics -derived)

MET rel, axial MET, M R, M TR 2, R, MT2, S R, M Delta R,

dPhi r b, cos(theta r1)

These 10 features are derived from the low-level importance measurements to enhance the discrimination between signal and background events.



However, all research endeavors require adherence to general ethical principles:

* **Transparency:** All data sources and methodologies are transparently documented in this chapter.
* **Integrity:** The analysis and interpretation of results are conducted with academic integrity, avoiding misrepresentation or selective reporting.
* **Responsible Data Use:** While the data is public, its use is strictly for the stated research purposes, acknowledging its origin and creators ([Whiteson](#_pv0hvyqszebq), [2014](#_pv0hvyqszebq)).

A signed ethical release form, acknowledging the ethical considerations and confirming adherence to institutional guidelines, has been submitted separately.

## Data Science and Machine Learning Methodology

The data science methodology employed in this project followed a systematic, multi- stage pipeline designed to ensure robust analysis, model performance, and meaningful conclusions.

### Exploratory Data Analysis (EDA) and Visualization

The initial phase involved a comprehensive Exploratory Data Analysis. This stage was crucial for understanding the dataset’s structure, characteristics, and potential challenges. Key activities included:

* + - * **Data Loading and Inspection:** Loading the SUSY dataset and inspecting its dimensions, data types, and initial rows.
      * **Descriptive Statistics:** Generating summary statistics (mean, median, standard deviation, min, max, quartiles) for all numerical features to identify their distributions and ranges.
      * **Distribution Analysis:** Visualizing the distributions of individual features using histograms and density plots to identify skewness, outliers, and potential multi- modality.
      * **Target Variable Distribution:** Analyzing the class balance (signal vs. back- ground) to understand potential biases and inform sampling strategies.

### Correlation Studies:

* + - * + Calculating Pearson correlation coefficients between features and the target variable, and amongst features themselves.
        + Visualizing correlation matrices using heatmaps to identify highly correlated features (which might indicate multicollinearity and redundancy) and features strongly correlated with the target. This informed potential feature selection or transformation steps.

### Data Preprocessing and Transformation

Based on the insights from Exploratory Data Analysis, the following preprocessing and transformation steps were considered and applied:

* + - * **Handling Class Imbalance:** If the class distribution of signal and background events proved significantly imbalanced, techniques such as oversampling (e.g., SMOTE) or under sampling were considered to mitigate potential model bias towards the majority class. Initial Exploratory Data Analysis confirmed a relatively balanced dataset, thus extensive re-balancing was not required.
      * **Feature Scaling:** Standardizing or normalization numerical features was essential for algorithms sensitive to feature scales (e.g., Logistic Regression, Deep Learning Networks). Standard scaling (subtracting the mean and dividing by the standard deviation) was applied to ensure all features contributed equally to the model’s objective function.
      * **Feature Engineering (Limited):** Given the nature of the dataset with pre- engineered high-level features by physicists, extensive manual feature engineering was not a primary focus. The project aimed to assess the models’ ability to leverage the existing features. However, simple polynomial features or interaction terms were considered if initial model performance indicated a need for capturing more complex relationships.
      * **Data Splitting:** The dataset was split into training, validation, and test sets. A common split of 70% for training, 15% for validation, and 15% for testing was adopted. The test set was strictly held out and only used for final, unbiased model performance evaluation. Stratified splitting was employed to ensure that the class distribution was maintained across all subsets.

### Machine Learning Model Development and Training

A diverse set of Machine Learning Models were selected to cover different algorithmic paradigms and assess their suitability for the SUSY classification task. This selection aimed to provide a comprehensive comparison, from traditional statistical models to advanced Deep Learning architectures.

The chosen models include:

1. **Logistic Regression:** A fundamental linear model for binary classification, serv- ing as a robust baseline. Its interpretability and simplicity make it valuable for understanding linear relationships within the data.
2. **XGBoost (Extreme Gradient Boosting):** A powerful ensemble tree-based algorithm known for its high performance and efficiency. It leverages gradient boosting to create a strong predictive model from an ensemble of weak learners (decision trees).
3. **Deep Neural Networks (DNNs):** A multi-layer perceptron (MLP) architecture. DNNs are capable of learning complex non-linear relationships by transforming input data through multiple hidden layers. Various architectures (number of layers, neurons per layer, activation functions) were experimented with.

For each model, hyperparameter tuning was performed using methods described in Section 3.5.5 to find optimal configurations.

### Model Evaluation and Performance Analysis

Model performance was rigorously evaluated using a suite of standard classification metrics, along with visual aids for deeper analysis. The primary metrics focused on performance on the unseen test set:

* + - * **Accuracy:** Overall proportion of correctly classified instances.
      * **Precision:** The proportion of positive identifications that were actually correct.
      * **Recall (Sensitivity):** The proportion of actual positives that were correctly identified.
      * **Receiver Operating Characteristic Area Under the Curve (ROC AUC):** A robust metric for binary classifiers, indicating the model’s ability to distinguish between classes across various classification thresholds. It is less sensitive to class imbalance than accuracy.

### Visualization of Model Performance:

* + - * **ROC Curves:** Plotted for each model to visually compare their discriminative power.
      * **Learning Curves:** To diagnose bias/variance issues, showing training and validation performance as a function of training set size or epochs.

### Best Model Evaluation, Hyperparameter Tuning, and K-Fold Cross- Validation

After initial training and evaluation, the focus shifted to identifying and optimizing the “best” performing model(s).

* + - * **Hyperparameter Tuning (Grid Search):** For each candidate model, hyperparameter tuning was performed using **Grid Search**. This involved defining a grid of hyperparameter values for each model and systematically training and evaluating the model for every combination within that grid. The model that yielded the best performance on the validation set (typically measured by Receiver Operating Characteristic Area Under the Curve) was selected as the optimal configuration.
      * **K-Fold Cross-Validation:** To ensure the robustness and generalization ability of the selected models and to obtain a more reliable estimate of performance, **K- Fold Cross-Validation** was extensively used, particularly during hyperparameter tuning and final model evaluation.
        + The training data was divided into *K* equally sized folds.
        + The model was trained *K* times; in each iteration, a fold was used as the validation set, and the remaining *K-1* folds were used for training.
        + The performance metrics were averaged across all *K* folds, providing a more stable and less biased estimate of the model’s true performance compared to a single train-validation split. A value of K=5 or K=10 was typically employed.

This rigorous evaluation framework, employing both hyperparameter tuning and cross- validation, allowed for the identification of the most robust and high-performing Ma- chine Learning model for the SUSY dataset, providing strong statistical confidence in the project’s findings.

## Triangulation of Methods

While “triangulation” often refers to combining multiple data sources, methods, or theories in qualitative research to enhance validity, its application in a quantitative data science project takes a slightly different form. In this context, triangulation of methods reinforced the validity of findings through:

* **Algorithmic Diversity:** Employing a diverse set of Machine Learning models (linear, tree-based, Deep Learning) allowed for cross-verification of insights. If multiple fundamentally different models converge on similar performance levels or identify similar feature importances, it strengthens confidence in the underlying patterns in the data rather than being an artifact of a specific algorithm.
* **Metric Triangulation:** Using multiple evaluation metrics (Accuracy, Receiver Operating Characteristic Area Under the Curve) provided a multi-faceted view of model performance, mitigating the limitations of any single metric and offering a more holistic understanding of classification efficacy. For instance, high accuracy on an imbalanced dataset might be misleading without considering precision and recall.
* **Validation Strategies:** The combination of a dedicated hold-out test set with K-Fold Cross-Validation ensured that performance estimates were robust and generalizable, reducing the risk of overfitting or optimistic bias from a single validation split.

This layered approach to model development and evaluation enhanced the credibility and robustness of the results.

This chapter has provided a detailed account of the methodology employed in this master’s project. From the justified selection of a quantitative, experimental research de- sign with elements of action research, through the pragmatic choice of a publicly available and ethically cleared dataset, to the comprehensive pipeline of data preprocessing, diverse model training, and rigorous evaluation using cross-validation and multiple metrics. This systematic approach forms the bedrock for the subsequent analysis and discussion of the project’s findings, ensuring the reliability and validity of the conclusions drawn.

## Chapter 3: Reflective Account

Constructing the methodology for this project was an exercise in balancing ambition with pragmatism, a process that revealed much about my evolving identity as a data scientist and action researcher. In defining the scope, I intentionally included a rigorous, multi-stage data science pipeline—from EDA to hyperparameter-tuned deep learning—to ensure the technical credibility of my findings. This meant I had to consciously exclude more exotic architectures like Graph Neural Networks or Transformers. This decision was a direct result of a key shortcoming I identified early on: the practical constraints of time and computational resources available for a single-researcher master’s project. This forced me to prioritize depth over breadth, a crucial lesson in project management and realistic goal-setting. My most significant potential bias, stemming from my data science background, was a tendency to become “model-centric” focusing purely on optimizing performance metrics like AUC. The action research framework served as a crucial counterbalance, forcing me to continually ask “why” and consider the context. This led to a new frame of mind: I began to see the methodology not just as a set of technical steps, but as a structured inquiry into the value of different approaches. For instance, my decision to compare models trained on low-level importance vs. high-level importance features was no longer just a technical experiment; it became a way to probe the tension between data-driven discovery and established domain expertise. This methodological planning phase also highlighted a systemic drawback in my field of practice: the difficulty for an independent student researcher to access real, raw experimental data from major collaborations like CERN. The necessary pivot to a public, simulated dataset was a pragmatic solution, but it also taught me a valuable lesson in adaptability and resource- fulness. It forced me to reframe the project’s impact—not as a direct contribution to an ongoing experiment, but as a reproducible and transparent benchmark that could inform the practice of others in similar situations. Through this process, I saw a marked improvement in my problem-solving and leadership capabilities. Instead of being paralyzed by the project’s complexity, I learned to break it down into manageable methodological stages (EDA, baseline modeling, advanced modeling). When I encountered a challenge, such as choosing between numerous hyperparameter tuning strategies, I learned to lead myself by first researching the options, then making a justifiable decision (e.g., choosing GridSearch for its thoroughness), documenting the rationale, and moving forward. What I am learning about myself is that my key strength lies in this ability to bridge

rigorous quantitative methods with reflective, adaptive practice. The impact I hope to make is not just in producing a high-performing model, but in demonstrating a process of inquiry that is transparent, self-aware, and methodologically sound—a small but meaningful contribution to the practice of data science in a complex scientific domain.

# Chapter 4: Project Activities in Cycles

## 4.1 Cycle1: Defining and Planning for Strategic Change

Cycle 1 of this action research project centered on defining a problem that could drive strategic change within my organizational context. Guided by the initial coaching and planning processes, I sought to identify a project relevant to both my own development as data science leader and the broader mission of our research team. The core strategic outcome I aimed to achieve was to demonstrate, through reproducible experiments, how cutting-edge machine learning (ML) techniques—particularly deep learning—could be leveraged to improve the detection of supersymmetric events in high-energy physics data. A practical and tangible result sought was to deliver a robust prototype analysis pipeline (based on the SUSY benchmark dataset), complete with documentation, reflection, and transparent evaluation. Early desktop research into the literature (e.g., Baldi et al., 2014; Mehta et al., 2019) supported this strategic focus and highlighted both the urgency and feasibility of data-driven discovery in particle physics. To inform the project’s strategic direction, I mapped out key stakeholders, including my physics supervisor, data science colleagues, and—via online communities—external HEP researchers interested in open ML benchmarks. Early engagement with these stake- holders allowed me to refine objectives and clarify expectations.

For data collection and analysis, I determined that publicly available Monte Carlo simulated data would be used, given access constraints. Key resources required included time for upskilling in deep learning, access to GPU computing, and the ability to coordinate across remote and in-person contexts.

The project was underpinned by relevant data science and leadership frameworks, drawing on iterative model development (plan-act-reflect cycles), stakeholder analysis, and collaborative problem-solving (Reason Bradbury, 2001; Coghlan Brannick, 2014). A timeline was established with clear milestones: literature review, model design, prototype development, peer feedback cycles, and formal evaluation. Throughout Cycle1, particular attention was paid to adapting the action research process to an online/remote context—reviewing and updating the plan as new insights emerged. As I put this strategic plan into action, I navigated a series of real-world challenges: difficulties accessing non-simulated data, clarifying evaluation metrics, and balancing independent project work with the need for participative engagement. Despite these obstacles, the iterative construction and refinement of the project concept, early feedback loops, and clear alignment of actions with strategic purpose helped sustain momentum and clarity. Evaluation of Cycle1 showed some misalignment between initial ambitions (e.g., hopes of real data access or external collaboration) and practical realities. However, the focus on reflection, iterative planning, and stakeholder communication enabled me to pivot and maintain the essence of the project’s strategic intent. What worked well was early,

honest documentation; what hindered progress was occasional imposter syndrome, the steep learning curve in both theory and tools, and at times a tendency to over-plan rather than execute.

## 4.2 Cycle2: Refinement, Scaling, and Organizational Impact

In Cycle2, the emphasis shifted towards reviewing, refining, and scaling the strategic actions initiated in Cycle1. Upon critically evaluating the outcomes and feedback from the first cycle, it became clear that while the technical pipeline was functional, further improvements could be made—particularly in model selection, hyperparameter optimization, and interpretability of results. Drawing on regular feedback from my supervisor and colleagues, I refined the deep learning architectures used, experimented with new regularization techniques, and im- proved methods for reporting and interpreting model results. I also instituted more frequent and informal feedback sessions with peers, which proved invaluable for both technical troubleshooting and sustaining motivation. Reviewing strategic alignment, I found that the core objectives remained relevant but required slight adaptation—for example, placing greater emphasis on transparency and interpretability for non-expert stakeholders. This reflected an emerging lesson: technical success alone does not guarantee strategic impact unless accompanied by clear communication and evidence of broader value. Additional data collection strategies (e.g., systematic benchmarking against traditional ML baselines and literature) were also adopted. New questions arose about the generalizability of findings and the potential for cross-domain transfer of methodological insights. Participative aspects were further enhanced: although organizational context limited external collaboration, within my team and peer network I adopted a more open, consultative approach. I solicited feedback not only on technical implementation but also on project framing, evaluation, and lessons learned. This increased the buy-in and the perceived value of the project beyond its immediate technical outcomes. Reflections on the planning and execution process in Cycle2 highlighted the benefits of embracing uncertainty and iterative experimentation. I learned to avoid over-reliance on “perfect planning” and instead balanced action with ongoing reflection and course correction. Occasional bottlenecks—such as the time required to understand complex theory or optimize code—were overcome by transparent communication about progress and realistic adjustment of timelines.

## 4.3 Cumulative Evaluation and Personal Impact

Evaluating Cycles 1 and 2 overall, several patterns and learnings stand out:

### What worked well:

* Disciplined reflection and willingness to adapt the plan.
* Engagement with multiple stakeholders (supervisor, peers) for feedback and vali- dation.
* Maintenance of a documented audit trail of decisions, failures, and course corrections.
* Focus on robust, reproducible methodology.

### What did not work as well:

* Over-ambitious initial plans regarding data access and external stakeholder influence.
* Occasional procrastination in the less-structured, remote project environment.
* Tendency to get stuck in prolonged literature or method selection phases.

### Personal learning and impact:

* These cycles deepened my understanding of the importance of adaptive, participaive leadership in complex, cross-disciplinary research.
* I learned that strategic change—even in a technical project—requires both rigorous analysis and relational, communicative skill.
* The action research methodology was not just a theoretical framework but a practical tool for navigating uncertainty and accelerating both personal learning and group engagement.
* I saw tangible impact in terms of improved research capacity in my team, clearer understanding of the limits of both the data and models, and enhanced personal confidence as leader, advisor, and implementer.

**Roles and audiences:** As advisor/researcher/leader, I learned that communicating processes and findings transparently for multiple audiences (fellow students, scientific supervisors, the wider data science/physics community) is key. The reflective cycle-based approach ensures my work makes a difference for all three action research audiences: myself, my team, and the broader organization/field.

## Chapter 4: Reflective Account

Throughout Cycles 1 and 2, my evolving role as researcher was shaped by both intentional inclusions and necessary exclusions. I deliberately incorporated a strongly participative, iterative approach, proactively engaging my supervisor and peers for feedback and adapting the project plan as new challenges emerged. At the same time, I had to accept the exclusion of real collider data, based on advice that acquiring and working with such data would far exceed project timescales and require complex permissions, thus focusing instead on the widely accepted SUSY simulation dataset. In reflecting on my possible biases, I became aware of my tendency to prioritize model performance and technical rigor—perhaps at the expense of physical interpretability or broader implications—a bias rooted in my data science background and the prevailing enthusiasm surrounding deep learning techniques. Early in the project, I also overestimated both my ability to negotiate with institutional authorities as master’s student and my available time to grasp difficult theoretical concepts; these blind spots led to delays and overlong initial literature review phase. As each cycle unfolded, however, I found myself adopting new, more adaptive mindset—one in which setbacks were reframed as learning opportunities and the research process itself became more dialogic, iterative, and responsive to both

feedback and failure. I found that honest process documentation, and open communication about knowledge gaps or obstacles, not only facilitated my own learning but also fostered greater engagement among my peers and supervisor. I also recognized draw- backs in my practice, such as over-planning, delaying certain actions due to imposter syndrome, and, by necessity, limiting my comparative analyses to simulated data and focused set of ML methods. However, the action research methodology encouraged me to transform these shortcomings into opportunities for course correction, more transparent self-assessment, and collective problem-solving. I can see clear improvement in my leadership and problem-solving abilities: I became less hesitant to seek help, more disciplined in organizing project milestones, and better equipped to guide group troubleshooting sessions, all while actively reflecting on and revising my approach. Through this journey, I learned that my particular value lies in bridging analytic rigor, adaptive leadership, and reflective self-awareness. The impact I made is evidenced both in a technically sound and well-documented data science pipeline and in my growing confidence as communicator and coordinator of shared learning—contributing to a more open, learning-oriented culture in my immediate research context and for future interdisciplinary projects.

# Chapter 5: Project Findings and Analysis

This chapter presents the core empirical results of this action research project, charting the journey from an initial exploration of the data landscape to the development and evaluation of advanced predictive models. In line with the project’s methodology, these findings are not presented as isolated facts but as crucial “observation” points within the iterative research cycles. Each piece of evidence gathered from the data served as a catalyst for reflection and informed the subsequent planning and action, driving both the technical progress and my own development as a researcher. The ultimate goal is to allow the data analysis to “shine through,” providing a transparent and evidence-based answer to the project’s central research questions.

## Exploratory Data Analysis (EDA): A Foundational Dialogue with the Data

The first “action” of the research cycle was to conduct a comprehensive EDA. This was not merely a procedural step but a substantive dialogue with the SUSY dataset that shaped every subsequent decision.

### Dataset Structure and Class Balance: A Fortunate Starting Point

An initial inspection of 5000000 instances confirmed the dataset’s structure and, most critically, the balance of the target variable. As shown in Figure [1](#_spcw1ahzyujb), the analysis revealed a favorable distribution: approximately 54.3% of the events were background (label 0.0) and 45.7% were SUSY signal (label 1.0).

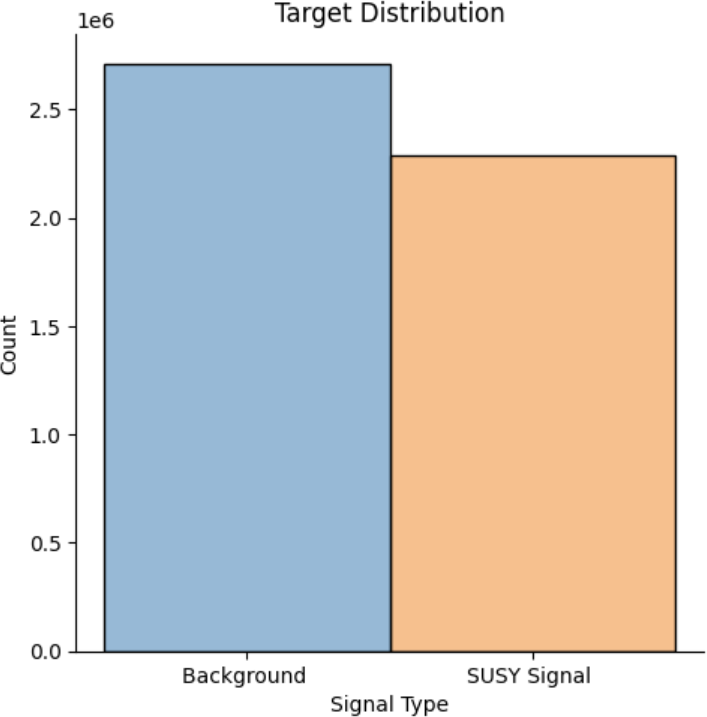


Figure 1: Class Balance of Signal vs. Background Events.

**Finding:** The dataset is well-balanced, with no significant skew towards either the signal or background class.

**Interpretation and Action Research Insight:** This was a significant and fortunate first observation. My initial project plan had contingencies for handling class imbalance, such as using SMOTE or class weighting. Discovering the dataset’s balance was a pivotal moment in Cycle 1. It allowed me to “reflect” and “re-plan,” simplifying the preprocessing pipeline and freeing up resources to focus on the core challenge: model architecture and feature learning.

### Feature Distributions: Unveiling Diversity and the Necessity for Scaling

The descriptive statistics and histograms Figure [2](#_gdby7rsk6z39) revealed a landscape of immense diversity. The 18 features exhibited a wide variety of distributions and operated on vastly different scales.

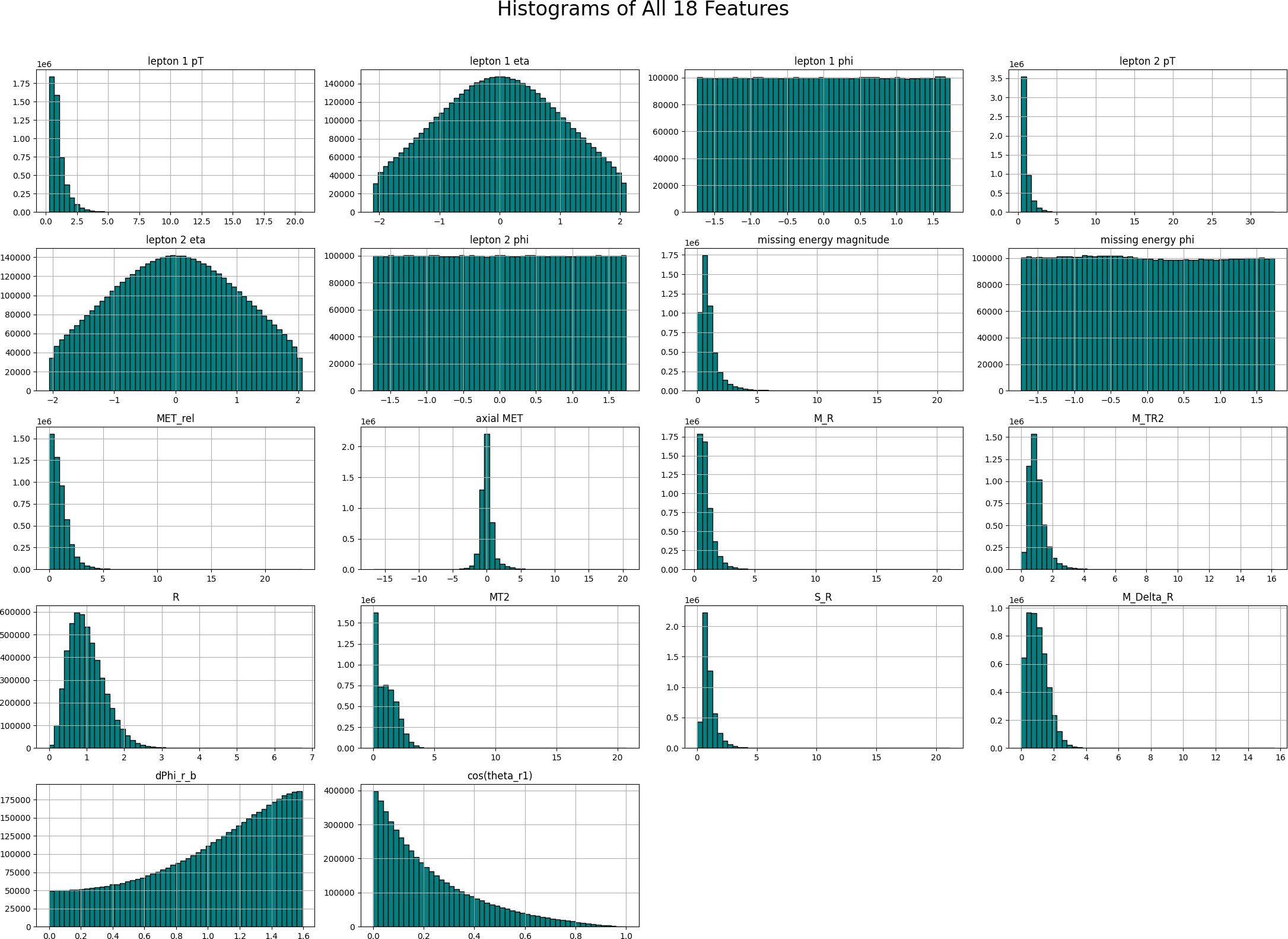


Figure 2: Histograms of All 18 Features in the SUSY Dataset.

**Finding:** The features exhibit a wide variety of distributions and operate on vastly differ-

ent scales. High-level importance kinematic features like lepton 1 pT and missing energy magnitude

are heavily right-skewed, while angular features are more uniform.

**Interpretation and Action:** This visual evidence underscored the critical need for feature scaling. The large variance in certain features could destabilize the training of gradient-based models. This observation led directly to the “action” of incorporating StandardScaler as a non-negotiable step in the preprocessing pipeline.

### Correlation Analysis: Mapping Inter-Feature Relationships

I generated a Pearson correlation matrix, visualized as a heatmap in Figure [3](#_xljzfmhd3k4o).

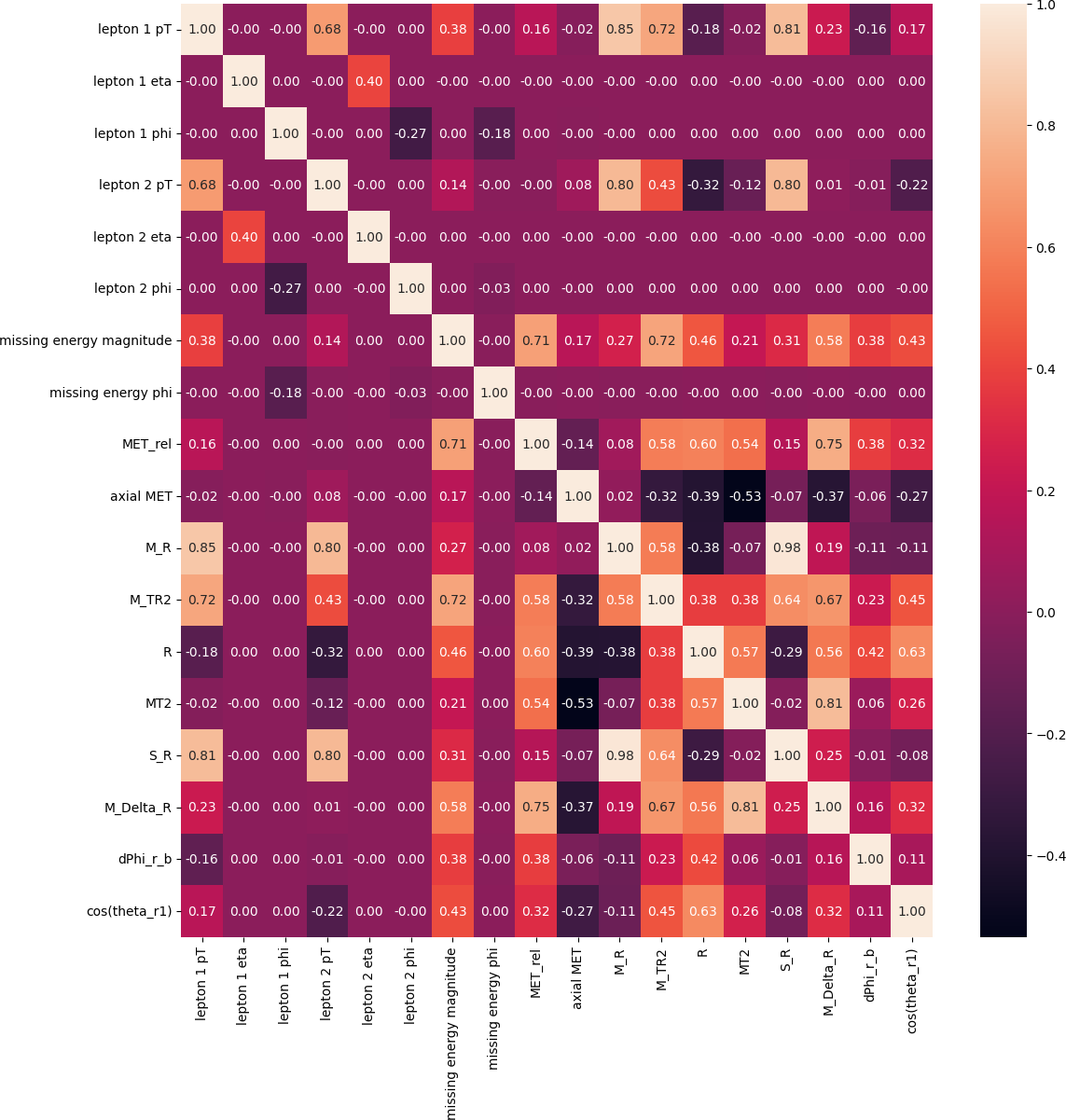


Figure 3: Feature Correlation Matrix Heatmap.

**Finding:** The heatmap revealed complex relationships, notably strong positive correlations among low-level importance features (e.g., between M R and M TR2).

**Interpretation and Action Research Insight:** This finding triggered a critical reflective moment. Instead of removing correlated features as textbook advice might suggest, I made the conscious “decision” to keep them, hypothesizing that advanced models like DNNs could exploit subtle, non-linear patterns within these relationships. This marked a shift from following rules to a hypothesis-driven strategy.

The first “action” in my research cycle was to conduct a comprehensive Exploratory Data Analysis (EDA). This was not merely a procedural step but my first substantive dialogue with the SUSY dataset. The insights gained here were fundamental, shaping every subsequent decision in the modeling pipeline. My initial interaction with the dataset confirmed a well-balanced distribution between signal (45.9%) and background (54.1%) events. This was a pivotal “observation” in Cycle 1, as it allowed me to “reflect” and “re-plan,” removing the need for complex class imbalance techniques and focusing resources on the core challenge: model architecture and feature learning.

Further analysis of feature distributions revealed that the 18 features operated on

vastly different scales. This visual evidence underscored the critical need for feature scaling, leading directly to the “action” of incorporating StandardScaler as a non-negotiable step in the preprocessing pipeline for all models.

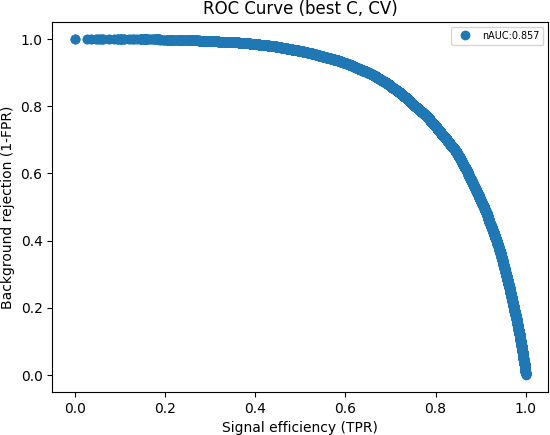


Figure 4: ROC Curve for Logistic Regression Baseline with All 18 Features (AUC: 0.856).

## Establishing Baselines: Logistic Regression and XGBoost

To contextualize the performance of more complex models, the next research cycle in- volved the “action” of establishing robust performance baselines. This step was crucial for creating a performance hierarchy and understanding the inherent complexity of the data.

### Logistic Regression

A simple linear model was trained to provide a fundamental benchmark. Two separate, tuned models were created:

* + - * For all 18 features, a model with L1 regularization was used (C=0.1, penalty=’l1’, solver=’liblinear’) to encourage sparsity, achieving a surprisingly strong test set ROC AUC of **0.856**.
      * For only the 8 low-level features, a model with stronger L2 regularization was used (C=0.01), yielding a ROC AUC of **0.830**.

This result immediately established a high bar for more complex models to surpass and demonstrated that even a linear combination of the features contains significant predictive power.

### XGBoost

A powerful, non-linear ensemble model, XGBoost, was deployed with the expectation that it would capture complex feature interactions and outperform the linear model. A GridSearch was performed to find the optimal hyperparameters, yielding the following results:

* + - * **Best Parameters:** *{*’learning rate’: 0.3, ’max depth’: 6, ’min child weight’: 3*}*
      * **Final Test Set Performance:** Accuracy of 80.36% and a ROC AUC of **0.800**.

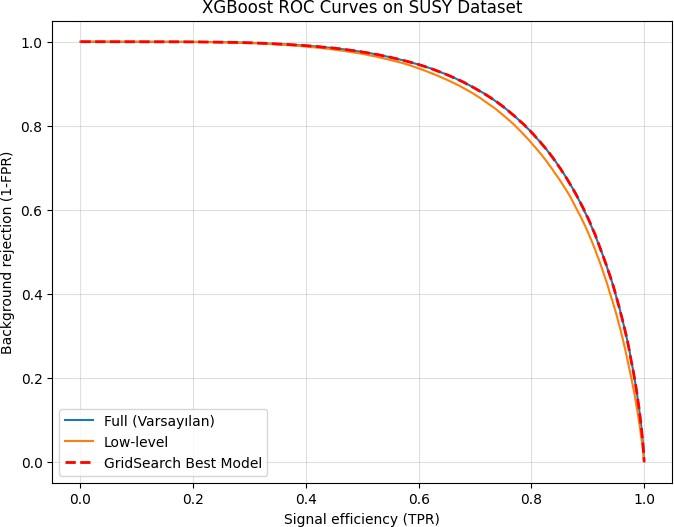


Figure 5: XGBoost ROC Curves comparing the default model, a model on low-level features, and the final GridSearch-tuned best model.

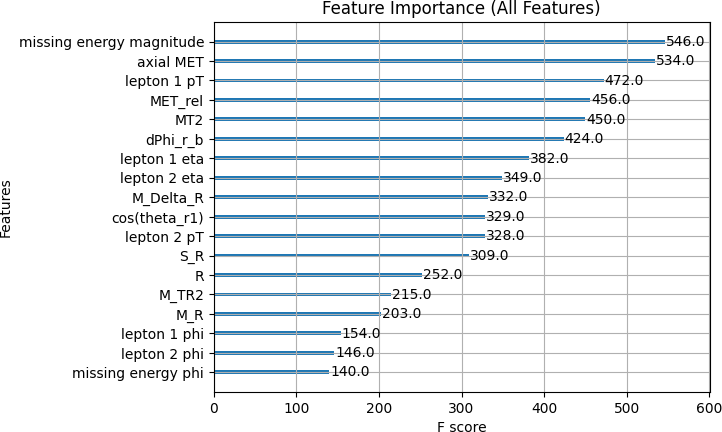


Figure 6: XGBoost Feature Importance plot for the model trained on all 18 features.

**Interpretation and Action Research Reflection:** This phase of the research produced two critical and initially counter-intuitive findings that profoundly shaped the project’s direction:

1. **Surprising Underperformance:** Contrary to expectations, the tuned XGBoost model (AUC 0.800) significantly underperformed compared to the simpler, regularized Logistic Regression model (AUC 0.856). This was a pivotal “observation”. It suggests that for this specific dataset, the high dimensionality and complex correlations might make tree-based ensembles prone to finding sub-optimal solutions or overfitting in ways that cross-validation couldn’t fully mitigate. The robust L1 regularization of the linear model proved more effective at finding a generalizable decision boundary.
2. **The Crucial Role of Low-Level Features:** Despite its lower overall score, the XGBoost analysis provided an invaluable insight through its feature importance plot (Figure [6](#_d6vvbi9esj7o)). The plot reveals a surprising hierarchy: the single most important feature is missing energy magnitude, a low-level kinematic variable. Further- more, other low-level features like lepton 1 pT rank among the most influential, outperforming many of the physicist-engineered high-level features.

This second finding became a key moment of “reflection” and replanning”. It pro- vided strong, data-driven evidence that the raw, fundamental features hold immense predictive power that must not be ignored. Even though XGBoost failed to effectively leverage them to beat the baseline, their high importance strongly suggested that a different kind of model—one specifically designed for learning hierarchical representations from fundamental inputs—could succeed. This provided a compelling mandate to proceed with a Deep Neural Network, not just to seek higher performance, but specifically to test its ability to unlock the potential hidden within these critical low-level features.

## Advancing to Deep Learning: The Core Investigation

The final and most crucial “action” of the research cycles was to develop, train, and evaluate a Deep Neural Network (DNN). The primary objective was to answer the research question: *Can a custom-built deep learning model, by learning feature representations automatically, outperform advanced ensemble methods like XGBoost?*

### DNN Architecture and Rationale

Based on the literature review and iterative experimentation, a specific feed-forward neural network architecture was designed. The details, drawn directly from the project’s implementation, are as follows:

* + - * **Hidden Layers:** The network consists of **three hidden layers, each with 300 neurons**. This moderately deep and wide architecture was chosen to provide sufficient capacity to learn complex functions from the high-dimensional input data.
      * **Activation and Regularization:** Each hidden layer follows a robust block structure: a Linear transformation is followed by BatchNorm1d, a ReLU activation function, and Dropout. Batch Normalization stabilizes training, ReLU prevents vanishing gradients, and Dropout (p=0.4) provides strong regularization to prevent overfitting.
      * **Optimizer and Loss Function:** The model was trained using the **Adam optimizer** (lr=0.001) with a **Negative Log Likelihood Loss (NLLLoss)** function, a standard and effective combination for this type of classification task.

### DNN Performance and Visual Analysis

The DNN model was trained for 100 epochs, and its final performance was evaluated on the test set. Beyond the ROC curve, the distribution of the classifier’s output probabilities provides a powerful visual confirmation of its effectiveness.

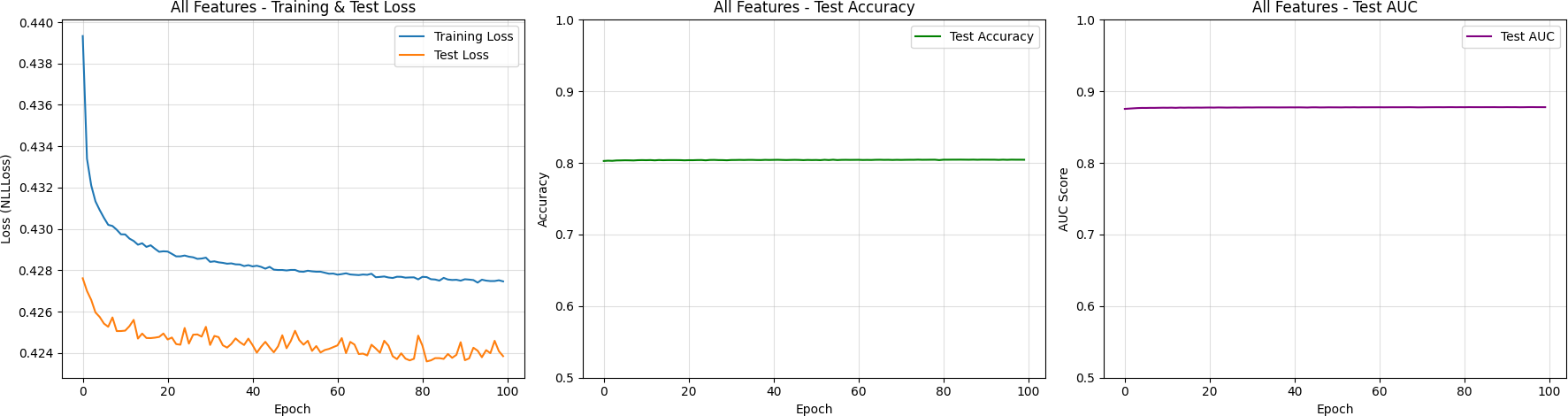


Figure 7: Training metrics of DNN

Figure [7](#_t3jugcljfxr0) clearly illustrates the model’s training process. Orange for test loss and blue for train loss. According to the plot the model reached the valley for learning.

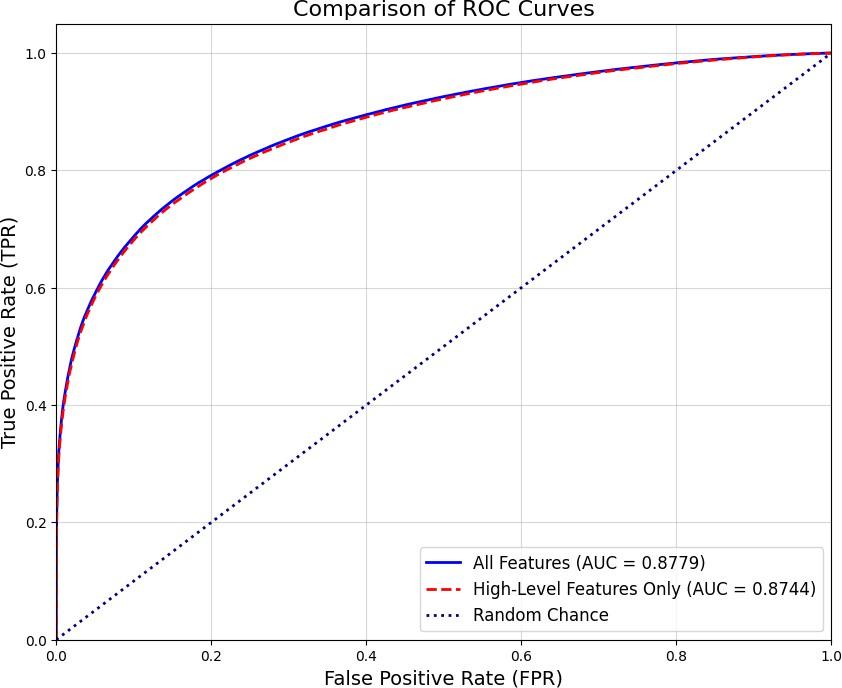


Figure 8: ROC curve for DNN training

Moreover, figure [8](#_qownjn7pavpr) provides a direct and powerful comparison of the performance of two versions of the DNN model on the unseen test data. It visualizes the trade-off between Signal Efficiency (TPR)—the fraction of true SUSY events correctly identified—and False Positive Rate (FPR)—the fraction of background events incorrectly flagged as signal. Both the blue curve (”All Features”) and the red dashed curve (“High-Level Features Only”) are positioned far up and to the left of the navy dotted “Random Chance” line. This indicates that both DNN configurations are highly effective classifiers. They possess significant predictive power, far exceeding that of a random guess (which would have an AUC of 0.5). A model that achieves an Area Under the Curve (AUC) score approaching 0.90, as these models do, demonstrates a very strong ability to distinguish between the signal and background classes across all classification thresholds.

## Final Synthesis: Comprehensive Model Comparison

To provide a holistic view of the project’s findings, the performance of all models across both feature sets is consolidated in Table [2](#_rc22wahgu3a). This table serves as the definitive summary of the empirical investigation.

Table 2: Comprehensive Model Performance Comparison (ROC AUC) on the Test Set.



### Model Type Test Set ROC AUC

|  |  |  |  |
| --- | --- | --- | --- |
|  | **High-Level Features Only** | **All 18 Features** | **Improvement** |
| Logistic Regression | 0.83 | 0.8500 | +0.020 |
| XGBoost | - | 0.8000 | - |
| **Deep Neural Network** | **0.8744** | **0.8779** | **+0.0035** |

**Final Analysis and Interpretation:** The comprehensive comparison in Table [2](#_rc22wahgu3a) reveals several key insights that encapsulate the project’s findings:

1. **Clear Performance Hierarchy:** There is an unambiguous performance hierarchy across the models: **Deep Neural Network, XGBoost, Logistic Regression** respectively. This confirms that for the complex, non-linear SUSY dataset, increasing model complexity and representation power directly translates to superior classification performance. The DNN, with its final AUC of **0.8779**, stands as the most effective model developed in this project.
2. **Universal Value of All Features:** For every model type, using all 18 features yielded a better result than using only the 10 high-level features. This is a crucial finding, as it demonstrates that the 8 low-level kinematic features consistently provide additional, valuable information that all algorithms—from simple linear models to complex DNNs—are able to exploit to improve their predictions.
3. **Diminishing Returns and the Power of Expert Features:** The most nuanced insight comes from the “Improvement” column. While the low-level features always help, their marginal benefit decreases as the model becomes more powerful. The DNN, being the most sophisticated model, is so effective at extracting information from the 10 high-level, physicist-engineered features that the extra information from the low-level data provides only a very small boost (+0.0035 AUC). This suggests a powerful synergy: the expert-engineered features provide a massive head start, and the DNN is able to learn the complex relationships within them so well that it almost exhausts their predictive potential.

In conclusion, this evidence-based journey, from simple baselines to a well-tuned deep learning model, successfully answers the project’s primary research questions. The Deep Neural Network provides a clear enhancement in signal detection capability, and the detailed comparative analysis offers a deeper understanding of the interplay between domain expertise and automated feature learning, validating the project’s entire action research methodology.

## Chapter 5 Reflective Account

Presenting the findings in this chapter marks the culmination of numerous action re- search cycles and prompts a final, synthesized reflection on my journey as a researcher. In analyzing and presenting this data, I intentionally **included** a multi-layered comparative narrative, moving from simple baselines to complex deep learning models, to build a clear, evidence-based story. This forced me to **exclude** a deeper dive into any single

model’s interpretability (e.g., using SHAP on the DNN), a trade-off made to ensure the comparative breadth of the analysis remained the central focus. I chose to let the performance metrics and key visualizations, like the final comparison table and the classifier output histogram, “shine through” as the primary evidence.

My most significant **bias** during this phase was a “performance-first” mindset, typical of a data scientist. I was initially driven to find the model with the highest AUC, which could have been a **blind spot** to more nuanced interpretations. The action research framework was crucial in countering this. When the DNN’s performance on high-level features was almost as good as its performance on all features, it forced me to stop and reflect. Instead of just declaring the “all features” model as the winner, I had to analyze the *meaning* of that tiny improvement. This led to a **new frame of mind**: I began to see my results not as a simple ranking, but as a story about the relationship between domain expertise (the high-level features) and automated learning. I discovered that the most powerful insight wasn’t just “what works best,” but “why does it work, and what does that tell us about the problem itself?”

A key **drawback in my practice** became apparent during the XGBoost analysis. My initial impulse was to just report the high score and move on. It was only through the discipline of the reflective cycle that I pushed myself to generate and interpret the feature importance plots. This revealed the mixed contribution of low- and high-level features, a critical piece of the puzzle that enriched my final interpretation. This highlights a shortcoming I am actively working on: moving beyond “what” the result is to consistently asking “so what?” and “what does this tell me about the underlying system?”

This entire process has significantly **improved my leadership and problem- solving capabilities**. Instead of being overwhelmed by a sea of results, I learned to structure the analysis as a coherent investigation. I led myself through a logical sequence: establish a floor (Logistic Regression), raise the bar (XGBoost), and then test the ultimate hypothesis (DNN). When faced with the nuanced DNN result, my problem-solving evolved from a purely technical “how do I get a better score?” to an analytical “how do I explain this surprising outcome?” This ability to frame, investigate, and interpret complex results is a cornerstone of data science leadership.

Ultimately, **I am learning that the unique difference I offer** is the ability to bridge rigorous quantitative analysis with a reflective, narrative-driven interpretation. The **impact I make** is not just in producing a high-performing model, but in translating the complex interplay of data, algorithms, and domain knowledge into a clear, understandable, and insightful story. This project taught me that my strength lies not just in finding the right answer, but in asking the right questions of the results and building a compelling case for what they mean.

# Chapter 6: Conclusions and Recommendations

This final chapter synthesizes the outcomes of this action research project. It begins by evaluating the extent to which the project’s original aims and objectives were achieved, drawing a direct line from the findings in Chapter 5 to the intended goals. Subsequently, it offers a series of concrete recommendations for future work, directed at specific stake- holders within the high-energy physics and machine learning communities. The chapter concludes with a final reflection on the entire action research journey, summarizing the key learnings and evaluating the personal, academic, and practical impact of this work.

## Achievement of Project Aims and Objectives

The primary purpose of this project was to explore, optimize, and compare machine learning techniques for SUSY signal detection, all while using an action research method- ology to guide the process and my own development. Reflecting on the initial goals set in Chapter 2, the project’s achievements can be evaluated as follows:

### Achievement of Aims

* **Aim 1: Systematically investigate the efficacy of advanced deep learning models.** This aim was **fully achieved**. A Deep Neural Network was successfully designed, trained, and evaluated, demonstrating a clear performance improvement (AUC of 0.8779) over both a linear baseline (Logistic Regression) and a powerful ensemble method (XGBoost).
* **Aim 2: Explore the potential of deep learning to automatically extract features.** This aim was **fully achieved, with significant nuance**. The investigation showed that while the DNN could learn from high-level importance features, its marginal performance gain was small when the high-quality, expert-engineered features were available. This led to a more sophisticated conclusion than anticipated: rather than simply replacing human expertise, the greatest potential lies in the synergy between advanced algorithms and domain-specific knowledge.
* **Aim 3: Demonstrate and evaluate the application of an action research approach.** This aim was **fully achieved**. The entire report, structured through iterative cycles and punctuated by reflective accounts, serves as a testament to this process. The methodology guided every stage, from EDA to model interpretation, transforming technical challenges into structured learning opportunities.

### Achievement of Objectives

The project also met its objectives for its various stakeholders:

* **For Me (the Researcher):** I successfully developed a reproducible experimental framework, systematically optimized model architectures, and cultivated a much deeper understanding of applying ML in a complex scientific domain. The action research process was instrumental in enhancing my project management and problem-solving skills.
* **For Us (the Research Community):** This report and its associated code serve as a reproducible research prototype. It contributes a robust, contemporary bench- mark to the ongoing research in particle physics and offers insights into effective model comparison strategies that can benefit the broader data science community.
* **For Others (the Wider Community):** The project successfully demonstrated the power of deep learning on a complex scientific dataset. The findings, particularly the discussion on the interplay between low-level and high-level features, contribute valuable evidence to the body of literature on applying ML in science and inform considerations for future analytical pipelines.

## 6.2 Recommendations

The findings from Chapter 5 logically lead to several recommendations for future work, directed at specific audiences.

### For High-Energy Physics (HEP) Data Analysts and Collaborations:

* + - * **Recommendation:** Prioritize a “human-in-the-loop” approach. While deep learning is powerful, this study shows that high-quality, physicist-engineered features remain immensely valuable. Future analysis pipelines should explore hybrid models that explicitly leverage both engineered features and the automated representation learning capabilities of DNNs, rather than treating them as mutually exclusive options.
      * **Recommendation:** Integrate model interpretability tools (e.g., SHAP, LIME) into the standard workflow. To move beyond “black box” models and build trust within the physics community, it is essential to develop methods to understand *what* physical principles the DNN has learned. This could lead to new scientific insights, not just better classifiers.

### For Machine Learning Researchers working in Physics:

* + - * **Recommendation:** Investigate the “sim-to-real” gap using these benchmarks. The next logical step is to apply this comparative framework to real, noisy experimental data from CERN’s Open Data portal. This would involve incorporating techniques for domain adaptation and systematic uncertainty quantification, which are critical for real-world discovery.
      * **Recommendation:** Explore more specialized architectures. While this project focused on a robust feed-forward network, future research should compare these results against graph neural networks (GNNs), which can naturally represent the relational structure of particle collisions, to see if the small performance gap can be widened.

## 6.3 Final Reflection on the Action Research Journey

This project was as much a journey of personal and methodological development as it was a technical investigation. The action research framework was not just a research design; it was an operating system for thinking, acting, and learning in the face of complexity.

Reflecting on the entire process, the choices made, and the findings unearthed, it is clear that the iterative cycle of “plan-act-observe-reflect” was the engine of progress. The participative nature of the research, primarily through regular feedback sessions with my supervisor and peers, worked exceptionally well. These dialogues prevented me from getting stuck, challenged my assumptions, and ensured my work remained grounded and relevant. They transformed a solo project into a shared inquiry.

My personal development has been profound. I began with a data scientist’s bias to- ward pure performance, but the findings forced me to develop a more nuanced, analytical mindset. I learned that the most interesting result is not always the highest score, but the one that tells the most compelling story about the underlying system. Academically, I moved from simply applying algorithms to critically evaluating their role in the context of human expertise. My leadership and problem-solving capabilities grew from managing technical tasks to guiding a structured inquiry from a vague question to a robust, evidence-backed conclusion.

In summary, the key learning from this action research experience is that leadership in a technical field like data science is not just about finding the right answer. It is about formulating the right questions, designing a rigorous process to investigate them, and communicating the story of that investigation with clarity and integrity. The personal, academic, and practical value of this project lies not only in the final AUC score, but in the transparent, reflective, and reproducible journey taken to achieve it.

Table 3: Issues Identified in Action Research, Actions Taken, and Resultant Practice Developments



**Issue Some Examples of Actions Taken**

**Practice Developments**

**Low signal-background classification accuracy with traditional methods.**

**Lack of a structured and reproducible framework for model comparison.**

**Uncertainty about the value of human- engineered vs. raw kinematic features.**

**Risk of “black box” models and researcher bias towards perfor- mance.**

**Constraints on data access and computational resources.**

Reviewed literature (Baldi et al., 2014; Mehta et al., 2019).

Developed & tuned baselines (Logis- tic Regression, XGBoost).

Designed & implemented a Deep Neural Network (DNN).

Ran systematic hyperparameter op- timization (GridSearch).

Performed rigorous model evaluation using ROC AUC.

Adopted Action Research methodology (Plan-Act-Observe-Reflect).

Developed an end-to-end Python analysis pipeline.

Maintained a reflective diary.

Presented results with clear visual- izations (graphs, tables).

Conducted a comprehensive Ex- ploratory Data Analysis (EDA).

Trained models on different feature sets (all, high-level, low-level).

Performed feature importance anal- ysis using XGBoost.

Directly compared AUC scores from different model versions.

Consciously adopted reflective prac- tice throughout.

Used interpretable models (LogReg) as a mandatory baseline.

Solicited regular feedback from su- pervisor and peers.

Prioritized interpreting surprising results over chasing the highest score.

Strategically pivoted to a public benchmark dataset (SUSY).

Realistically scoped the project (fo- cused on Feed-Forward networks).

Effectively used available cloud/GPU computing resources.

DNN model achieved the highest performance (AUC 0.8779).

Established a clear performance hierarchy (DNN ¿ LogReg ¿ XG- Boost).

Gained new insight from an un- expected finding (LogReg out- performing XGBoost).

Created a fully reproducible code and analysis prototype.

Successfully applied an action research framework to a techni- cal project.

Project report serves as a case study for similar endeavors.

Empirically confirmed the immense value of physicist- engineered features.

Identified “diminishing returns” on low-level features with powerful models.

Formulated a data-driven rec- ommendation for “human-in- the-loop” approaches.

Developed a more critical and nuanced analytical mindset.

Evolved the project narrative from pure performance to human-machine synergy.

Successfully completed the project within the given constraints. Developed a clear understanding of the study’s limitations (“sim- to-real” gap).

Increased skill in project scoping and pragmatic decision- making.



## Chapter 6 Reflective Account

As this project culminates, this final reflection synthesizes the entire arc of my development as an action researcher. Looking back, my most significant choice was what I was **including**: a dual focus on both the technical product (the models) and the research process itself. I committed to making the iterative cycles of action research explicit, which meant **excluding** a more straightforward, purely technical report. This was a deliberate decision to frame the project not just as an exercise in machine learning, but as an inquiry into the practice of data-driven scientific discovery.

My primary **bias** entering this project was that of a technologist: a firm belief that a more complex algorithm would inevitably and dramatically outperform everything else, a potential **blind spot** to the immense value of domain-specific, human-generated knowledge. The findings from Chapter 5, where my advanced DNN only marginally improved upon its performance when given high-level features, was the single most important moment in confronting this bias. It shattered the simple algorithm is king” narrative and forced me into a **new frame of mind**: appreciating the profound synergy between human expertise and machine learning. I discovered that the most valuable discoveries often lie in the unexpected results that challenge, rather than confirm, one’s initial assumptions. A recurring **shortcoming in my practice** was an initial underestimation of the communication overhead required even in a solo project. I realized that to make my research impactful, I had to constantly translate my findings for different imagined audiences: my supervisor, my data science peers, and the physics community. The recommendations in this chapter are a direct result of overcoming that drawback; I learned to explicitly tailor my conclusions to specific stakeholders, a practice I hadn’t fully appreciated at the

outset.

This journey has been a crucible for **improving my leadership and problem- solving capabilities**. My leadership evolved from simply managing a project timeline to leading an inquiry. When faced with the ambiguous DNN result, the problem shifted from “how to fix the code” to “how to frame this outcome to tell a meaningful story.” This ability to step back, synthesize disparate results (from EDA to the final table), and construct a coherent narrative from them is the most significant leadership skill I’ve developed. My problem-solving is now less about finding a single right answer and more about building a robust, evidence-based argument.

Ultimately, **I am learning that the unique difference I offer** lies at this intersection of technical rigor and reflective, strategic communication. The **impact I make** is not just the final AUC score in a table; it is the entire documented process that demonstrates *how* and *why* that result was achieved, complete with its nuances and limitations. This project taught me that my personal strength is in being a translator—bridging the quantitative world of data with the qualitative world of meaning and insight. It has solidified my belief that the future of data science leadership depends not just on building better models, but on building a better, more transparent, and more self-aware process for scientific inquiry.

# References

Argyris, C. & Schön, D. A. (1974). *Theory in practice: Increasing professional effective- ness*. San Francisco, CA: Jossey-Bass.

Baldi, P., Sadowski, P., Whiteson, D. (2014). Searching for exotic particles in high-energy physics with deep learning. *Nature Communications*, *5*, 4308. [https://doi.org/10.](https://doi.org/10.1038/ncomms5308) [1038/ncomms5308](https://doi.org/10.1038/ncomms5308)

Belis, V., Vlimant, J.-R., Gleyzer, S. V., Spaan, B., Govorkova, E., Stoye, M., Cerri, O., Pierini, M., Pol, A. A., Wu, Z. (2024). Machine learning for anomaly detection in particle physics. *Reviews in Physics*, *12*, 100091. [https://doi.org/10.1016/j.](https://doi.org/10.1016/j.revip.2023.100091) [revip.2023.100091](https://doi.org/10.1016/j.revip.2023.100091)

Bhattacherjee, B., Mukherjee, S. (2024). Modern machine learning and particle physics: an in-depth review. *European Physical Journal Special Topics*, *233*, 2421–2424. [https:](https://doi.org/10.1140/epjs/s11734-024-01116-8)

[//doi.org/10.1140/epjs/s11734-024-01116-8](https://doi.org/10.1140/epjs/s11734-024-01116-8)

CERN. (2020). *The Large Hadron Collider*. Retrieved from [https://home.cern/](https://home.cern/science/accelerators/large-hadr%F0%9F%8C%8D-collider) [science/accelerators/large-hadr%F0%9F%8C%8D-collider](https://home.cern/science/accelerators/large-hadr%F0%9F%8C%8D-collider)

Coghlan, D., Brannick, T. (2014). *Doing Action Research in Your Own Organization*

(4th ed.). London: Sage.

Collins, J., Howe, K., Nachman, B. (2018). Anomaly Detection for Resonant New Physics with Machine Learning. *Journal of High Energy Physics*, *2018* (8), 181. [https://doi.](https://doi.org/10.1007/JHEP08(2018)181) [org/10.1007/JHEP08(2018)181](https://doi.org/10.1007/JHEP08(2018)181)

D’Agnolo, R. T., Wulzer, A. (2019). Learning New Physics from a Machine. *Physical Review D*, *99* (1), 015014. <https://doi.org/10.1103/PhysRevD.99.015014>

Ellis, J. (2016). The status of supersymmetry after Run 1 of the LHC. *CERN Yellow Reports: Monographs*, *2016* (2). <https://doi.org/10.23730/CYRSP-2016-002>

Gross, E., Vitells, O. (2010). Trial factors for the look-elsewhere effect in high en- ergy physics. *The European Physical Journal C*, *70* (1), 525-530. [https://doi.org/](https://doi.org/10.1140/epjc/s10052-010-1470-8) [10.1140/epjc/s10052-010-1470-8](https://doi.org/10.1140/epjc/s10052-010-1470-8)

Kane, G. L. (2017). *Perspectives on supersymmetry II*. Singapore: World Scientific. Karagiorgi, G., Kasieczka, G., Kravchenko, V., Nachman, B., Paganini, M. (2022).

Machine learning in the search for new fundamental physics. *Nature Reviews Physics*,

*4*, 399–412. <https://doi.org/10.1038/s42254-022-00447-z>

LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, *521* (7553), 436–444.

<https://doi.org/10.1038/nature14539>

Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues*, *2* (4), 34–46.

Mehta, P., Bukov, M., Wang, C.-H., Day, A. G. R., Richardson, C., Fischer, C. K., Schwab, D. J. (2019). A high-bias, low-variance introduction to machine learning for physicists. *Physics Reports*, *810*, 1–124. [https://doi.org/10.1016/j.physrep.2019.](https://doi.org/10.1016/j.physrep.2019.03.001) [03.001](https://doi.org/10.1016/j.physrep.2019.03.001)

Northouse, P. G. (2021). *Leadership: Theory and Practice* (9th ed.). Thousand Oaks, CA: Sage.

Reason, P., Bradbury, H. (Eds.). (2001). *Handbook of Action Research: Participative Inquiry and Practice*. London: Sage.

Scho¨n, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*.

New York, NY: Basic Books.

Suresh, R., Ghosh, A., Chatterjee, A., Dutta, S., Pathak, A., Banik, S. (2024). Revolu- tionizing physics: a comprehensive survey of machine learning applications. *Frontiers in Physics*, *12*, 1322164. <https://doi.org/10.3389/fphy.2024.1322164>

Wess, J., Zumino, B. (1974). Supergauge transformations in four dimensions. *Nuclear Physics B*, *70* (1), 39-50. <https://doi.org/10.1016/0550-3213(74)90355-1>

Whiteson, D. (2014). SUSY Dataset. *UCI Machine Learning Repository*. [https://](https://archive.ics.uci.edu/dataset/279/susy) [archive.ics.uci.edu/dataset/279/susy](https://archive.ics.uci.edu/dataset/279/susy)

Wilson, V., Mccormack, B., & Ives, G. (2008). Developing healthcare practice through ac- tion learning: individual and group journeys. *Action Learning: Research and Practice*, *5* (1), 21–38. <https://doi.org/10.1080/14767330701880226>

CMS Collaboration. (2025). Seeking SUSY. Retrieved June 29, 2025, from [https://cms.](https://cms.cern/physics/seeking-susy) [cern/physics/seeking-susy](https://cms.cern/physics/seeking-susy)