

An Educational Resource for Particle Identification (PID) in Experimental Particle Physics Data

Richard Dube

Honors Advisor

Richard Jones

University Scholar Committee

Richard Jones

Diego Valente

Myron Minn-Thu Aye

Abstract

Although particle physics research is typically reserved for advanced undergraduates, algebra-based physics students can readily develop the foundational skills necessary to conduct particle physics research. Despite this, there is a shortage of educational resources that introduce particle physics to algebra-based physics students. This project, Particle Identification Playground, is a collection of Python-based activities that teach students about several fundamental topics in experimental particle physics research. Through these activities, students will learn about the common detectors used in particle physics, how they work, and how we can use these detectors to identify particles. Students are able to interact with 3D models of real particle detectors, use applets to visualize and analyze particle physics data, and create and interpret plots commonly seen in published papers. These activities provide students with the rare opportunity to engage with particle physics research as early as high school.

1 Introduction

The traditional path into particle physics begins with some form of affiliation with a university. Undergraduate students will approach a professor who conducts particle physics research, and will join their lab group to learn foundational topics of particle physics. Students will simultaneously develop auxiliary skills required for particle physics research, such as coding. However, this path is inaccessible to a significant population. For example, students that do not attend a research university may find it difficult to find a mentor in the field, which may prevent them from exploring the field of particle physics. This issue is particularly severe for high school students, who may require more mentorship than most professors are willing or able to provide.

In recent years, several resources have been developed which enhance the average student's ability to engage with experimental particle physics. For example, in 2014, the European Center for Nuclear Research (CERN), which oversees the Large Hadron Collider (LHC), announced its open data portal, which has since released over two petabytes of experimental data for the public to analyze [1]. This reduces the burden on the curious student to find a mentor at a university to find data to analyze, but it does not equip students with the skills required to analyze the released data.

Other projects, such as Particle Physics Playground¹ and QuarkNet², provide activities that teach students about the data analysis methods used in experimental particle physics. These activities mainly focus on a stage of data analysis in particle physics called event reconstruction. The goal of event reconstruction is to figure out what interactions occurred and what particles decayed to produce the particles that we eventually detected in our apparatuses. However, the detectors used in particle physics experiments don't directly provide information about what particles we detected. Instead, we need to analyze the signals left in the detector to infer the type of particle that left said signal.

This is the goal of the particle identification process: to identify particles based on the signals they leave in particle detectors. However, there are no educational resources that teach budding particle physicists how to identify particles in experimental data, which prevents curious students from developing the skills to analyze experimental data, even if said data is publicly available.

This project created Particle Identification Playground³, a website that hosts a collection of jupyter notebooks that introduce students with no particle physics experience to the particle identification process. These activities only assume users have basic knowledge of algebra, algebra-based physics, and python, which makes them accessible for students as early as high school. These activities are also hosted on Google Colaboratory, which allows students to complete these activities without needing to download or install anything on their computer. With these activities, there is now a complete set of resources that enable students to thoroughly analyze publicly available particle physics data without requiring mentorship or affiliation with a research university.

¹<https://sites.google.com/siena.edu/particle-physics-playground/home>

²<https://quarknet.org/content/home-page>

³<https://duberii.github.io/pid-playground/>

2 Activities

This section contains a brief summary of each of the activities hosted on Particle Identification Playground as of April 2024. For a complete, updated list of activities you can visit Particle Identification Playground directly at <https://duberii.github.io/pid-playground/>. You can also click on the headers of each section to be brought to the corresponding activity.

2.1 Introductory Activities

2.1.1 Introduction to Vectors

“Introduction to Vectors” introduces vectors in the context of physics. The notebook uses a python library called Plotly to create interactive 3d plots of vectors, which allows students to see how the properties of a vector change as they vary its components. These interactive plots are also used to help students see the effect of multiplying a vector by a scalar, which they will use in later notebooks. Figure 1 shows an example of a plot generated by Plotly. The user is able to click and drag to rotate the axes and the vector, which will help students visualize what a vector would look like if plotted in 3D space. This notebook is vital to ensure that high school students have worked with vectors before seeing them used in later notebooks.

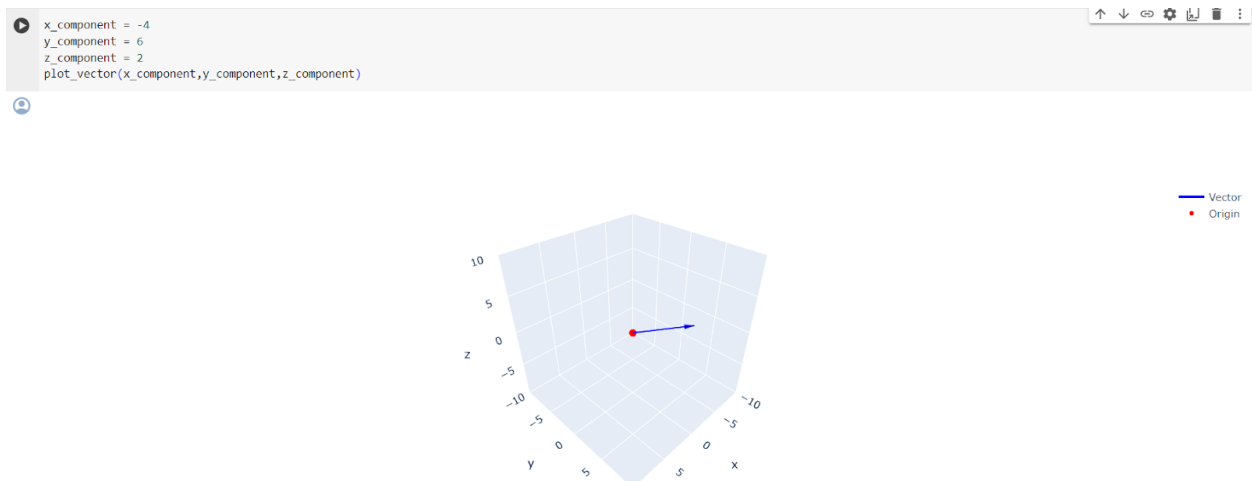


Figure 1: The solution to a practice problem in the “Introduction to Vectors” notebook. The 3D plot shown can be rotated to allow the student to see the plot from different angles in 3D space.

2.1.2 Introduction to Special Relativity

“Introduction to Special Relativity” introduces students to several relativistic quantities that are frequently used in particle physics, as well as the unit system used in particle physics. This notebook provides students with experience interpreting and manipulating

the definitions of relativistic momentum, relativistic energy, and invariant mass, which are quantities that are used to characterize particles that are detected in the GlueX detector.

2.1.3 Introduction to Pandas

“Introduction to Pandas” teaches students how to use a python package named Pandas, which provides a convenient way to store large quantities of data. Although pandas is not the industry standard for particle physics data analysis, it is a useful tool that students will be able to use in a wide variety of applications, including the machine learning applications that will be discussed in later activities. This activity introduces students to the dataframe object and introduces ways to add columns, calculate columns based on other columns, access rows, and iterate over rows, which allows for the rapid processing of large volumes of data.

2.1.4 Introduction to the Standard Model

“Introduction to the Standard Model” is where students start learning about particle physics. The student is introduced to foundational vocabulary in particle physics, which enables them to classify particles based on their properties. Students are taught how to identify bosons, fermions, hadrons, baryons, and mesons based on properties provided to them in a Pandas dataframe.

2.1.5 Introduction to ROOT

“Introduction to ROOT”, teaches students about the most widely used data analysis software in particle physics. In this notebook, students are introduced to the mission of particle identification, and then they use ROOT to produce a one-dimensional histogram that plots the masses of commonly detected particles, as shown in Figure 2. They then are asked to produce similar histograms for datasets with decreased energy and momentum accuracy, which helps them associate the width of a peak in a histogram with measurement errors and statistical fluctuations.

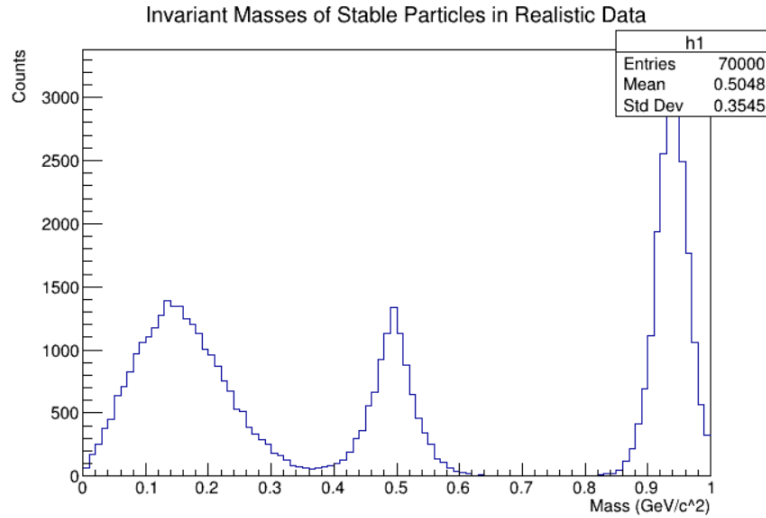


Figure 2: An example of a one-dimensional histogram that students are asked to produce in the “Introduction to ROOT” notebook. This histogram plots the invariant masses of 11 stable particles and is used to show that experimental error reduces our ability to distinguish particles.

2.1.6 2D Histograms in ROOT

In “2D Histograms in ROOT”, students are taught how they can plot two-dimensional data in a two-dimensional histogram. Students are taught how to modify the procedure introduced in the previous notebook to create a 2D histogram of energy vs. momentum, which is shown in Figure 3. Students are also taught that statistical fluctuations and experimental error present as “blurriness” or greater spread in 2D histograms.

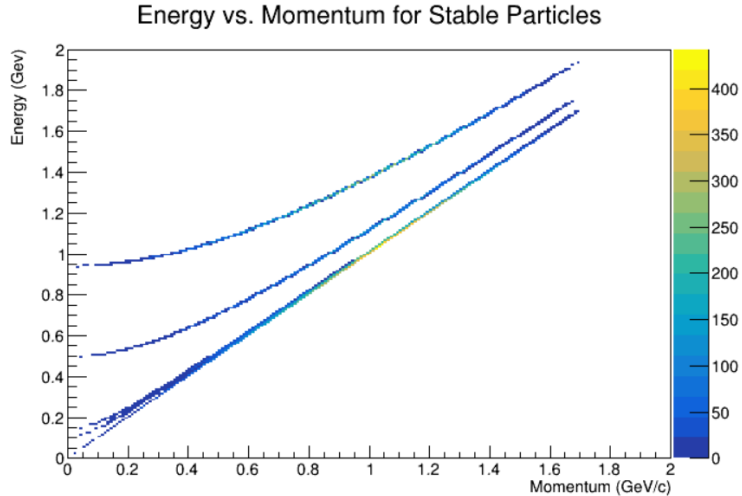


Figure 3: An example of a two-dimensional histogram that students are asked to produce in the “2D Histograms in ROOT” notebook. This histogram plots the energy vs the momentum of particles with no smearing or experimental error. Each stripe is associated with a different stable particle.

2.2 Particle Identification Methods

2.2.1 Drift Chambers and Tracks

“Drift Chambers and Tracks” introduces students to drift chambers, which are among the most common types of detectors used in particle physics experiments. In particular, students are introduced to the GlueX Central Drift Chamber (CDC). This notebook covers the geometry of the CDC, the working mechanism of drift chambers, and the track fitting algorithm, which together enable accurate measurements of the momenta of charged particles. Figures 4 and 5 show examples of interactive plots that students interact with as part of this notebook.

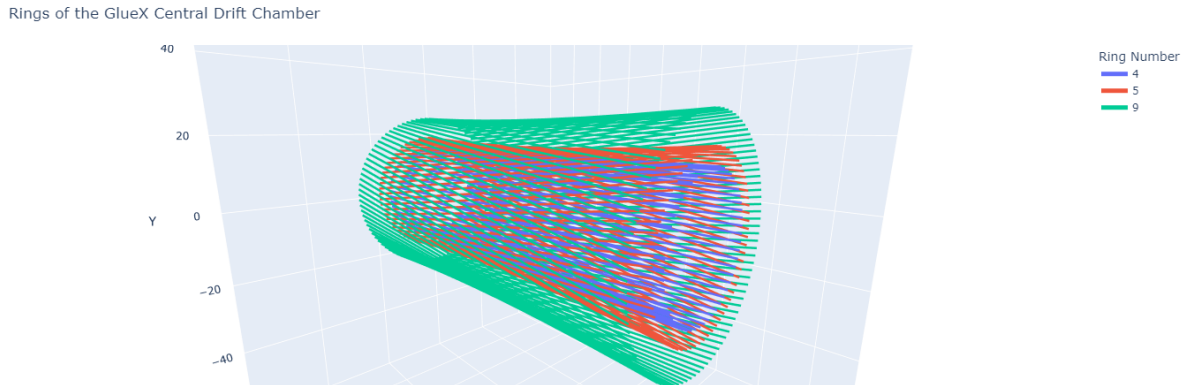


Figure 4: A 3D model of the wires of the GlueX Central Drift Chamber (CDC). Students are able to use sliders to change which rings they look at, which allows them to modify the visual complexity of the model.

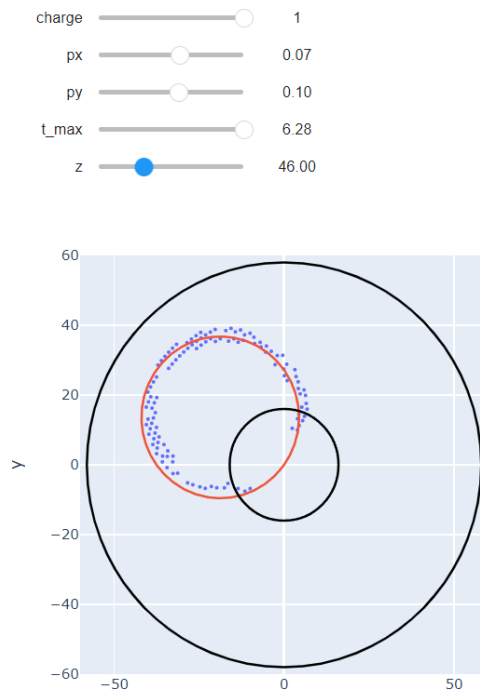


Figure 5: The track fitting applet in the “Drift Chambers and Tracks” notebook. Students adjust the ‘z’ slider to line up the hit points, then change the values of ‘px’ and ‘py’ to adjust the curvature of the track (shown in red). Running this cell multiple times will produce different hit patterns to fit, which allows students to get as much practice as they would like.

2.2.2 Ionization Energy Loss

In the “Ionization Energy Loss” notebook, students are taught about how raw detector signals in drift chambers can be used to calculate a quantity called ionization energy loss per unit length, which can be used to identify charged particles. Students are then walked through the process of creating an ionization energy loss histogram, such as the one seen in Figure 6. Finally, students are taught how they can use this plot to identify particles by using selection criteria known as cuts. The precise cuts are found by the student using an interactive applet that uses sliders to adjust the cuts in real time.

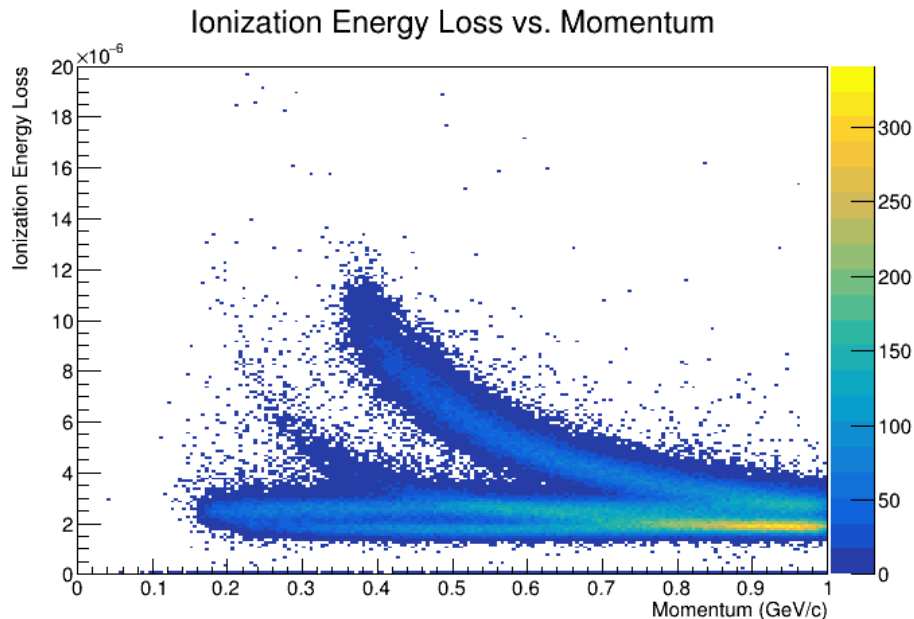


Figure 6: A 2D histogram of rate of ionization energy loss (dE/dx) and momentum (p). Each of the stripes in the histogram represents a distinct charged particle, and students learn how to use these stripes to identify particles.

2.2.3 Time of Flight Detectors

Time of flight detectors are another common type of detector used in particle physics experiments. In the “Time of Flight Detectors” notebook, students see how the GlueX Forward Time of Flight detector works. Furthermore, students are asked to derive the relationship between the quantities that are measured by the time of flight detector, momentum, and mass, which shows them how the time of flight detector can be used to classify particles using 2D histograms, such as the one shown in Figure 7. Finally, students place cuts on the time of flight data, which allows them to classify particles in a variety of GlueX detectors.

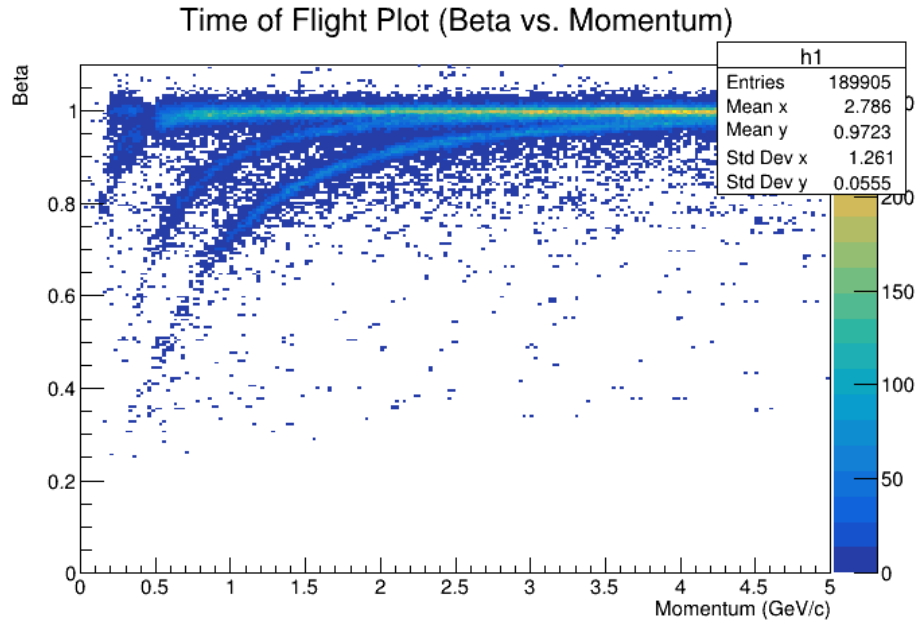


Figure 7: A 2D histogram students are asked to create in the “Time of Flight Detectors” notebook. The relativistic quantity β , which can be calculated based on the time of flight, is plotted against the momentum (p), which gives rise to distinct stripes corresponding to different particles that were detected.

2.2.4 Cherenkov Detectors

When an object moves faster than the speed of sound in air, it creates a sonic boom. When a charged particle moves faster than the speed of light in a material, it emits Cherenkov radiation. In the “Cherenkov Detectors” activity, students see how Cherenkov radiation can be used to identify particles. Students use the equations of special relativity to predict the angle at which a particle will emit Cherenkov radiation, as well as the minimum momentum that a particle has to have to emit Cherenkov radiation. Finally, students learn how the GlueX Detection of Internally Reflected Cherenkov radiation (DIRC) detector works, and uses it to identify particles.

2.2.5 Calorimeters and Neutral PID

All of the aforementioned particle identification methods only work on charged particles. However, there are several neutral particles that we frequently detect in particle physics. In “Calorimeters and Neutral PID”, students are introduced to calorimeters, which are the only GlueX detectors that can detect neutral particles. Students explore the geometry of the GlueX Barrel Calorimeter (BCal) through 3D models, such as the one shown in Figure 8, and visualize the complex interactions of particles in the BCal, as shown in Figure 9. Students then conclude that identifying neutral particles manually is extremely difficult, which provides motivation for the use of machine learning.

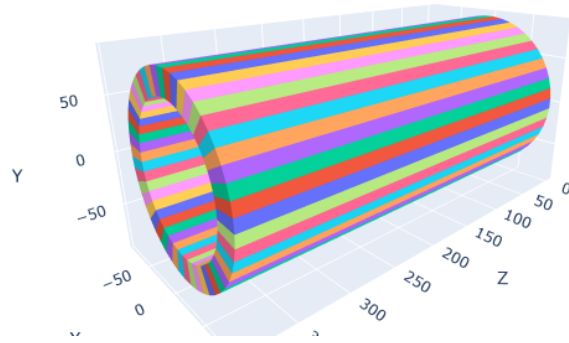


Figure 8: An interactive 3D model of the Barrel Calorimeter (BCal) of the GlueX experiment. Each of the blocks is a trapezoidal module, which are stacked together to form the cylindrical BCal.

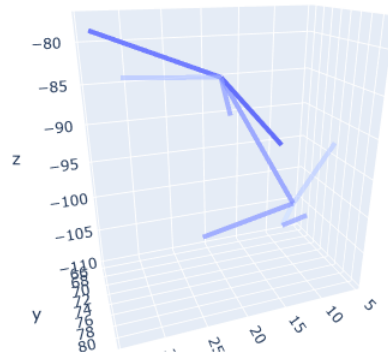


Figure 9: A 3D visualization of an interaction with the GlueX Barrel Calorimeter (BCal). The branching pattern comes from interactions between the material of the BCal and the particle, which produces a shower of secondary particles that branch out from the path of the particle.

2.3 Advanced Topics

2.3.1 Working with Experimental Data

In previous activities, data has been provided to students in the form of a dataframe and has been slightly simplified. In “Working with Experimental Data”, students are introduced to the structure of the files that are produced by the detector and the type of information that we have to work with. In particular, students are introduced to hypotheses, which are calculations of the properties of a particle based on an assumption of a specific particle type.

Multiple hypotheses are generated from each signal, so the role of particle identification is to figure out which hypothesis accurately describes the signal. Students restructure the functions they wrote for previous activities to be compatible with the hypothesis-based format used in experimental data.

2.3.2 Confusion Matrices

In the “Confusion Matrices” notebook, students are introduced to confusion matrices, which provide information about the accuracy of particle identification methods and identify particles that are particularly difficult to distinguish. Creating confusion matrices allows a direct comparison of the performances of two different particle identification methods. In this activity, students are taught how to generate and interpret confusion matrices, such as the one shown in Figure 10.

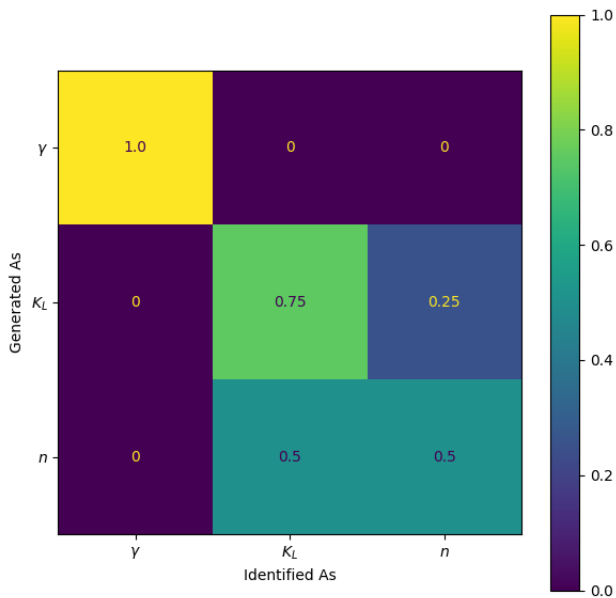


Figure 10: An example confusion matrix that is generated in the “Confusion Matrices” notebook. Rows represent the true identity of the particle, while columns represent the predicted identity using the particle identification method. The diagonal elements of the confusion matrix represent the accuracies for each particle type.

2.3.3 Comparing PID Methods

In “Comparing PID Methods”, students draw conclusions about the power of the previously discussed particle identification methods. They explore the performance of the particle identification functions they wrote on a variety of datasets and use the resulting confusion matrices to draw conclusions about the ability of each particle identification method to identify particles under a variety of conditions.

2.3.4 Machine Learning Methods

Although cuts-based particle identification is effective and intuitive, it is not the only approach to particle identification. Machine learning can provide an even more accurate particle classifier by finding more nuanced relationships between the features we use for cuts-based particle identification. In “Machine Learning Methods,” students learn to create and train a Scikit-learn boosted decision trees model and use the trained model to identify particles. Furthermore, students compare cuts-based particle identification with machine-learning-based particle identification and describe the advantages and disadvantages of both.

2.3.5 Bias in Machine Learning

“Bias in Machine Learning” is designed to emphasize the importance of minimizing sources of bias that may be introduced to machine learning models during the training or evaluation process. Students learn to define and identify common issues in machine learning, such as bias and overfitting. Feature importance is introduced as a method of identifying over-reliance on a particular feature of the model, which may be a sign that we have introduced some type of bias. Students are then introduced to several examples of biased models and are tasked with removing the primary source of bias from the training and evaluation processes.

3 Future Plans

3.1 Integrating Experimental Data

Although the current set of activities provides students with the skills to analyze real experimental data, Particle Identification Playground only provides simulation data. Simulation data is particularly easy to work with because we always know the exact processes and reactions that are being simulated, which we can use as an “answer key” for our particle identification methods. In experimental data, particle identification is the means of investigating the reactions that produced the particles we detected, which makes it much harder to create a labeled dataset. However, it is important that students are able to explore real data, so integrating experimental data is a priority.

3.2 Event Reconstruction

Although particle identification is a necessary stage of data analysis, it is only useful as preparation for event reconstruction. Therefore, a future activity is planned that will have students use their particle identification methods to identify particles then store the identified particles in a format that can be used for reconstruction. This way, students can make use of the activities provided by other resources such as Particle Physics Playground and QuarkNet.

3.3 Packaging Analysis Tools

Although many of the applets used in the activities of Particle Identification Playground exist for educational purposes and may not be useful for analysis of experimental data, some applets and functions provide a simplified wrapper for otherwise difficult to generate visualizations. Because of the potential of these wrappers to be useful in data analysis, packaging and publishing these functions may provide a simplified toolset for particle physics data analysis in Python. Although these functions were already developed as they were needed in the various activities, modifying the functions to be flexible enough to be useful in a variety of analysis contexts and documenting these functions is a nontrivial task.

4 Acknowledgements

This project would not be possible without the support of the University Scholar advisory committee, which included Richard Jones, Diego Valente, and Myron Minn-Thu-Aye.

I have worked with Richard Jones since my junior year of high school. Over the past six years, he has helped me hone my knowledge of particle physics, coding, and machine learning, which have been invaluable skills in my physics career. I am fortunate to have such a versatile mentor, and I aspire to follow in his steps as a brilliant physicist.

Diego Valente was instrumental in the development of the activities of Particle Identification Playground. His expertise in physics education has not only enhanced the quality of my work but has also broadened my perspective on effective teaching methodologies. With his mentorship, I have expanded my perspective on the experiences of physics students, which will guide my career as a physics educator.

Myron Minn-Thu Aye's support and kindness have been a constant source of encouragement, and have motivated me to pursue a masters degree in mathematics. Although the pandemic caused the first year of my college experience to unfold entirely online, Myron was a professor and advisor I knew I could always turn to, which helped ensure my success at UConn.

I would also like to express my appreciation to Erin Scanlon and Matt Guthrie for their mentorship in the field of physics education. Their expertise, commitment to accessibility, and constant faith in me have helped me grow as both a physicist and an educator.

Special thanks to Fabiana Cardetti for her guidance during my master's in mathematics. Her support and encouragement paved the way for me to explore opportunities in education through the TRANSCEND Fellowship. I am grateful for her mentorship and the opportunities she has provided me to foster inclusivity in the field of education.

Each of these individuals has played a significant role in shaping my academic and personal growth, and this project would not be possible without their support.

References

- [1] *CERN makes public first data of LHC experiments*. Nov. 2014. URL: <https://home.web.cern.ch/news/news/accelerators/cern-makes-public-first-data-lhc-experiments>.