Hybrid Ensemble Approach for Particle Track Reconstruction and Classification in High-Energy Physics

June 16, 2025

Abstract

Particle track reconstruction is a critical task in high-energy physics (HEP), where detectors capture vast amounts of raw data from particle collisions. This paper presents a hybrid ensemble approach that integrates multiple machine learning techniques to reconstruct particle tracks, classify particle types, and predict kinematic properties such as momentum and energy. Our method leverages clustering algorithms (HDBSCAN, K-Means, GMM, Agglomerative Clustering), convolutional neural networks (CNNs) for spatial feature extraction, long short-term memory networks (LSTMs) for sequential trajectory modeling, and fully connected regression networks for parameter estimation. We evaluate our approach on Monte Carlo-simulated detector events from CERN's Open Data Portal, achieving robust clustering and track reconstruction while identifying areas for improvement in predicting extreme kinematic values. This framework offers a scalable solution for HEP research, with potential for real-time implementation in large-scale experiments.

1 Introduction

Particle accelerators, such as the Large Hadron Collider (LHC), generate enormous volumes of high-dimensional data from subatomic particle collisions at relativistic speeds. The resulting detector signals must be processed to reconstruct particle tracks, identify particle species, and extract physical parameters like energy and momentum. Traditional methods, such as Kalman filters and combinatorial track-finding algorithms, are computationally expensive and struggle in dense collision environments [**?**].

Recent advancements in machine learning (ML) and deep learning (DL) offer new opportunities to enhance track reconstruction accuracy and efficiency. However, single-model approaches often fail to generalize across diverse detector conditions, necessitating ensemble-based strategies. This work proposes a hybrid ensemble model combining clustering, deep learning, and sequence modeling to improve track reconstruction and particle classification.

2 Problem Definition

2.1 Particle Track Reconstruction

Given raw detector hits $\mathbf{H} = \{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n\}$, the goal is to reconstruct particle tracks $\mathbf{T} = \{\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_m\}$, where each track corresponds to a distinct particle trajectory through the detector.

2.2 Particle Classification and Parameter Estimation

For each reconstructed track $\mathbf{t}_i$, the objective is to classify the particle type $P(\mathbf{t}_i) \in \{\text{electron, muon, pion, etc.}\}$ and estimate its kinematic properties: momentum (p), energy (E), and charge (q).

3 Methodology

Our approach uses a multi-stage pipeline: 1. **Unsupervised Clustering**: HDBSCAN, K-Means, GMM, and Agglomerative Clustering group detector hits into candidate tracks. 2. **Feature Extraction**: CNNs process spatial hit data. 3. **Trajectory Modeling**: LSTMs model temporal dependencies. 4. **Regression**: Fully connected networks predict momentum and energy.

4 Experimental Setup and Evaluation

4.1 Dataset

We use Monte Carlo-simulated detector events from CERN's Open Data Portal, split 80/20 for training and testing.

4.2 Evaluation Metrics

The table below lists the metrics used to evaluate our approach:

Table 1: Evaluation Metrics

Task	Metric	Description
Clustering	Silhouette Score	Measures track formation accuracy
Classification	F1 Score	Evaluates particle type classification
Regression	Mean Squared Error (MSE)	Assesses momentum and energy estimation
Overall	Reconstruction Efficiency	Fraction of correctly identified tracks

5 Results

5.1 Clustering Performance

We visualize raw and clustered particle hits:

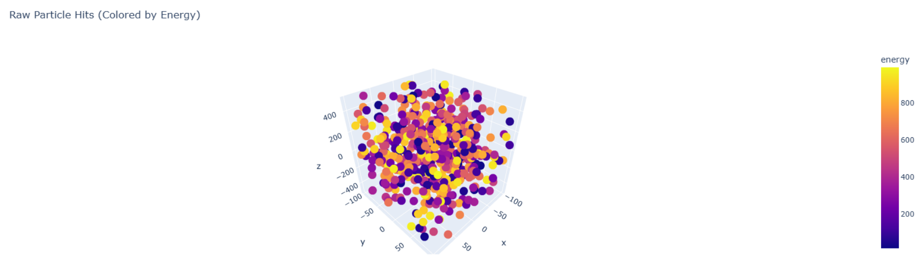

Figure 1: Raw Particle Hits (Colored by Energy).

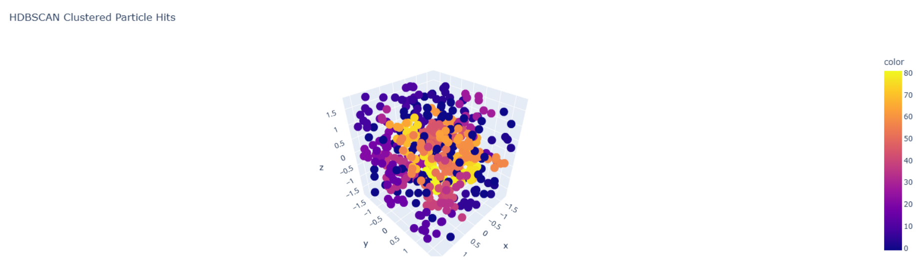

Figure 2: HDBSCAN Clustered Particle Hits (81 clusters).

5.2 Model Training

Training and validation losses are shown below:

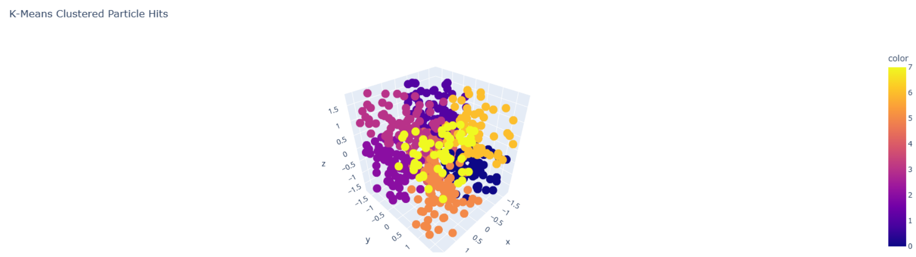

Figure 3: K-Means Clustered Particle Hits (8 clusters).

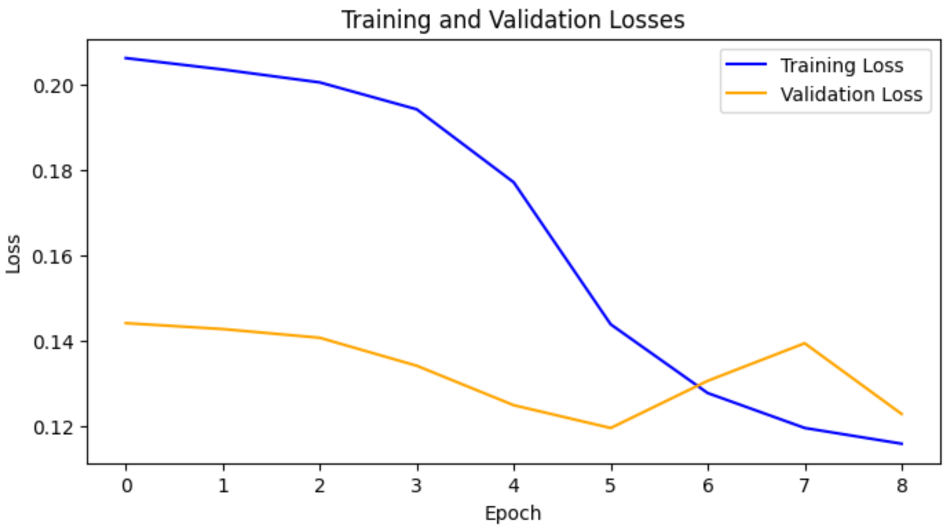

Figure 4: Training and Validation Losses over 8 Epochs.

5.3 Track Reconstruction

Reconstructed tracks are visualized in 3D:

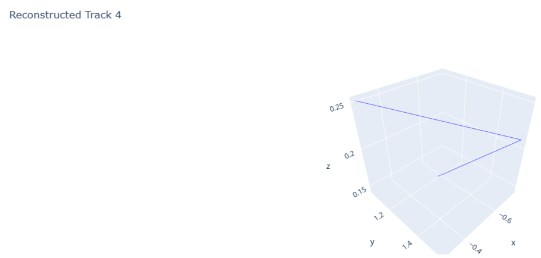

Figure 5: Reconstructed Track 3.

6 Conclusion

Our hybrid ensemble framework successfully reconstructs particle tracks and estimates kinematic properties, offering a scalable solution for HEP experiments.