

# Study on Hadronic Energy Reconstruction of High-Granularity Calorimeter Based on Machine Learning

Borong Peng<sup>1</sup>, Hongbin Diao<sup>1</sup>, and Yunlong Zhang<sup>\*1</sup>

<sup>1</sup>University of Science and Technology of China

October 15, 2026

## Abstract

To meet the demand for high-precision hadron energy measurements at the Circular Electron Positron Collider (CEPC), this study investigates machine-learning-based energy reconstruction for positive pions in the analog hadron calorimeter (AHCAL). Conventional energy-sum methods do not fully exploit the spatial structure of particle showers and are therefore limited in their ability to correct event-by-event fluctuations in hadronic showers. To address this limitation, three energy regression models are developed: a multilayer perceptron (MLP), a convolutional neural network (CNN), and ParticleNet. The MLP uses high-level statistical features, including the layer-wise deposited energy, number of hits, and average energy deposition. The CNN represents the detector response as a three-dimensional tensor and combines two-dimensional convolutions, attention mechanisms, and a bidirectional long short-term memory network to model the transverse morphology and longitudinal development of hadronic showers. ParticleNet represents the calorimeter hits as a point cloud and employs dynamic graph convolutions to integrate spatial, energy, and timing information. The models are evaluated using both Monte Carlo (MC) simulations and CERN beam-test data. The results show that the three models significantly outperform the conventional method under both test conditions, achieving

---

<sup>\*</sup>Corresponding author. Email: zhangyl@ihep.ac.cn

average relative improvements in energy resolution of more than 20% and improved response linearity. Among the models considered, ParticleNet provides the best overall performance. These results demonstrate that machine learning models can effectively exploit the high-granularity shower information recorded by the AHCAL and exhibit a certain degree of transferability from simulation to experimental data. This study provides a useful reference for particle-flow reconstruction and calorimeter performance optimisation at the CEPC.

**Keywords:** Circular Electron Positron Collider; analog hadron calorimeter; positive pion; energy reconstruction; machine learning

# 1 Introduction

The Circular Electron Positron Collider (CEPC) is a next-generation large-scale electron-positron collider proposed by the Chinese high-energy physics community [1, 2]. It is designed to enable precision measurements of the properties of the Higgs boson and electroweak interactions while providing an important experimental platform for searches for physics beyond the Standard Model. The CEPC detector adopts the particle-flow algorithm for event reconstruction, in which the momenta of charged particles are measured by the tracking system, photon energies are measured by the electromagnetic calorimeter, and neutral-hadron energies are measured by the hadron calorimeter [1, 3]. Within this reconstruction framework, the hadron calorimeter is required not only to measure hadron energies accurately but also to provide sufficiently fine spatial information to separate showers produced by nearby particles. Its energy resolution and shower-separation capability therefore have a direct impact on the jet energy resolution and boson mass resolution [1, 3].

The analog hadron calorimeter (AHCAL) is one of the principal candidate technologies for the CEPC hadron calorimeter [4]. It is a sampling calorimeter composed of alternating absorber plates and plastic-scintillator active layers. Its highly granular structure enables three-dimensional imaging of particle showers and provides detailed information for particle-flow reconstruction and shower energy measurements. To validate the performance of this technology, the CEPC AHCAL prototype has undergone several beam-test campaigns at CERN, during which large samples of muons, electrons, and charged pions

were collected [4, 5].

Energy reconstruction for hadronic showers is considerably more challenging than that for electromagnetic showers. When an incident hadron undergoes inelastic nuclear interactions with the absorber material, various secondary particles are produced, giving rise to coupled electromagnetic and non-electromagnetic shower components. The electromagnetic fraction, nuclear breakup processes, production and propagation of low-energy neutrons, and invisible energy losses exhibit large event-by-event fluctuations [6–8]. Consequently, the relationship between the visible energy in the calorimeter and the incident particle energy is nonlinear and depends on the detailed shower topology. In addition, the shower starting position and longitudinal development fluctuate substantially from event to event, and part of the shower may extend beyond the effective calorimeter volume, resulting in longitudinal or transverse energy leakage. Detector effects, including readout thresholds, silicon photomultiplier (SiPM) nonlinearity, electronic noise, and channel-to-channel response nonuniformity, further affect the reconstructed energy [5]. Conventional hadron energy reconstruction methods generally sum the visible energies of all hit cells and apply a common calibration factor to correct the average calorimeter response [8]. Although this approach is physically transparent and computationally simple, it corrects only the average energy scale of the sample and cannot adequately account for invisible energy, non-compensation, and energy leakage on an event-by-event basis according to the shower topology [7]. Two events with similar total visible energies may therefore correspond to different incident energies because of differences in their internal shower composition and spatial development. This is one of the principal limitations on further improvements in hadron energy reconstruction.

Conventional calorimeter energy reconstruction is typically based on summing the visible energies recorded by all active cells and then applying a global energy calibration or software-compensation procedure to estimate the incident particle energy [6]. These methods are physically interpretable but make limited use of the spatial correlations and local shower structures recorded by highly granular calorimeters. With the rapid development of high-granularity calorimetry and deep-learning techniques, machine learning methods have been increasingly applied to calorimeter particle identification, shower reconstruction, and incident-energy regression [9–11]. Machine learning models can learn the complex nonlinear relationship between detector responses and incident particle en-

ergies from large simulated or experimental data samples [12, 13]. By jointly exploiting the spatial distribution, longitudinal development, local energy deposition, and timing information of particle showers, these models can correct event-by-event fluctuations that are difficult to describe using conventional energy-sum methods [13]. Machine learning therefore provides a promising approach to improving the energy resolution and response linearity of highly granular calorimeters. Previous studies have demonstrated that machine learning methods can effectively exploit the spatial and timing information provided by high-granularity calorimeters and improve the energy resolution for both electromagnetic and hadronic showers [12]. However, differences between simulated and experimental data may arise from detector response, calibration, noise, and operating conditions. Model performance must therefore be evaluated not only using simulated samples but also with beam-test data to assess its generalisation to realistic detector conditions [6].

This work investigates energy reconstruction for positive pions in the AHCAL using three machine learning models: a multilayer perceptron (MLP), a convolutional neural network (CNN), and ParticleNet. The models are trained using Monte Carlo simulated samples and evaluated with independent Monte Carlo test samples and CERN beam-test data. Particular attention is paid to their performance when transferred from simulation to experimental detector data.

The remainder of this paper is organised as follows. Section 2 describes the AHCAL prototype, the CERN beam tests, the Monte Carlo simulation and detector digitisation, and the construction of the datasets. Section 3 presents the conventional energy-sum method, the three machine-learning-based energy reconstruction approaches, the model training strategies, and the performance metrics. Section 4 reports the energy reconstruction results obtained with Monte Carlo simulations and beam-test data. Finally, Section 5 summarises the main findings and discusses prospects for future work.

## 2 AHCAL Prototype and Data Samples

This section describes the CEPC analog hadron calorimeter prototype, the CERN beam-test data, and the Monte Carlo simulated samples used in this study. The overall structure and signal readout of the AHCAL prototype are first introduced, followed by a description of the beam-test setup and the positive-pion data used in the analysis. The detector

geometry, particle generation, and detector-response digitisation adopted in the Monte Carlo simulation are then presented. Finally, the construction of the training, validation, single-energy test, and beam-test samples, together with the corresponding data preprocessing procedures, is described.

## 2.1 General Structure of the AHCAL Prototype

The CEPC analog hadron calorimeter (AHCAL) is a highly granular sampling calorimeter composed of alternating iron absorber plates and plastic-scintillator active layers [4]. Incident particles interact with the iron absorbers and produce secondary particles, giving rise to showers that deposit visible energy in the plastic scintillators. By sampling the shower repeatedly along the calorimeter depth, the AHCAL records its longitudinal evolution from initiation and development to attenuation. In the transverse plane, each active layer is segmented into a large number of independent readout cells, providing detailed measurements of the transverse shower morphology and local energy distribution [4].

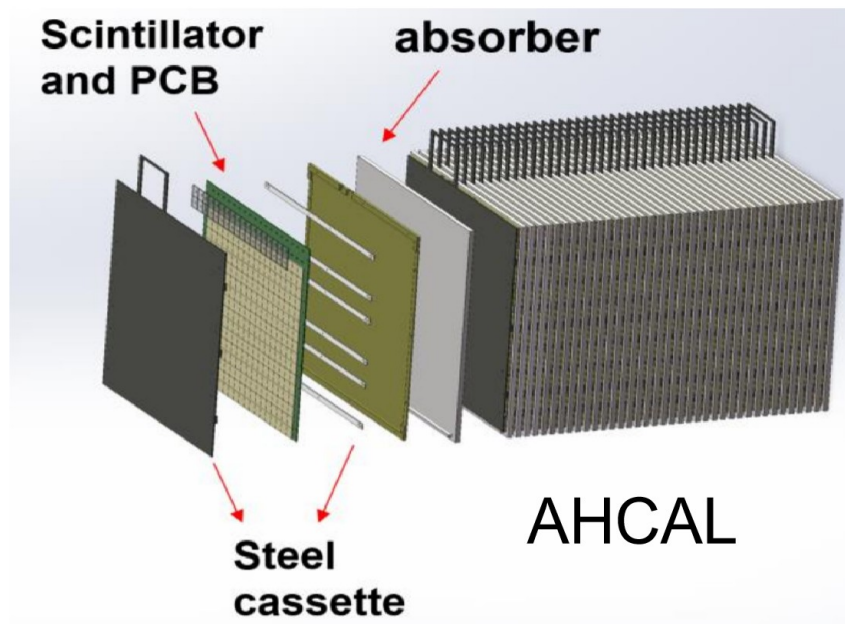


Figure 1: Schematic view of the CEPC AHCAL prototype.

The technological AHCAL prototype used in this study consists of 40 sampling layers and has a total depth of approximately 4.6 nuclear interaction lengths [1]. Each sampling layer contains a 20-mm-thick iron absorber plate and one plastic-scintillator active layer. The active area of each layer is  $720 \times 720 \text{ mm}^2$  and is segmented into an  $18 \times 18$  array of plastic-scintillator tiles, each with dimensions of  $40 \times 40 \times 3 \text{ mm}^3$  [1]. Each layer

therefore contains 324 independent readout cells, giving a total of 12,960 active cells and corresponding readout channels in the full prototype. This configuration records the energy-deposition pattern of each event on a regular three-dimensional grid of  $40 \times 18 \times 18$ . Each plastic-scintillator tile is coupled to a silicon photomultiplier (SiPM). Energy deposited by a particle in the scintillator produces scintillation light, which is collected by the SiPM and converted into an electrical signal. The signal is subsequently amplified, shaped, and digitised by the embedded front-end electronics. Each active layer is equipped with nine readout chips, with each chip serving 36 channels, thereby providing readout for all 324 scintillator tiles in the layer [1]. The data acquisition system uses a parallel readout architecture to accommodate the relatively high event rates encountered during the beam tests.

The high granularity of the prototype is important in both the longitudinal and transverse directions. Along the longitudinal direction, the 40 sampling layers, separated by approximately 20 mm, provide detailed measurements of the development of hadronic showers from their starting point to their attenuation. In the transverse plane, the  $40 \times 40 \text{ mm}^2$  cell size allows substructures within a shower to be resolved and helps distinguish the dense electromagnetic core from the more diffuse hadronic component. This fine spatial resolution provides the essential data foundation for the machine-learning-based energy regression studied in this work. By exploiting the detailed shower topology, machine learning models can learn and correct topology-dependent energy losses that are difficult to compensate using conventional energy-sum methods, thereby improving the energy reconstruction performance.

## 2.2 CERN Beam Tests

To evaluate the detector performance of the AHCAL prototype, the CEPC calorimeter group conducted three beam-test campaigns at CERN. The tests were performed at the H8 and H2 beam lines of the Super Proton Synchrotron (SPS) and the T9 beam line of the Proton Synchrotron (PS), with each campaign lasting approximately two weeks. Large samples of muon, electron, and charged-pion events were collected, together with electronics calibration and temperature-monitoring data. These data were used to study the channel response, operational stability, particle-identification capability, and energy reconstruction performance of the prototype. Figure 2 shows the prototype installed in

the beam-test area. The detector was mounted on a movable platform, and the black structure corresponds to the AHCAL prototype [1].

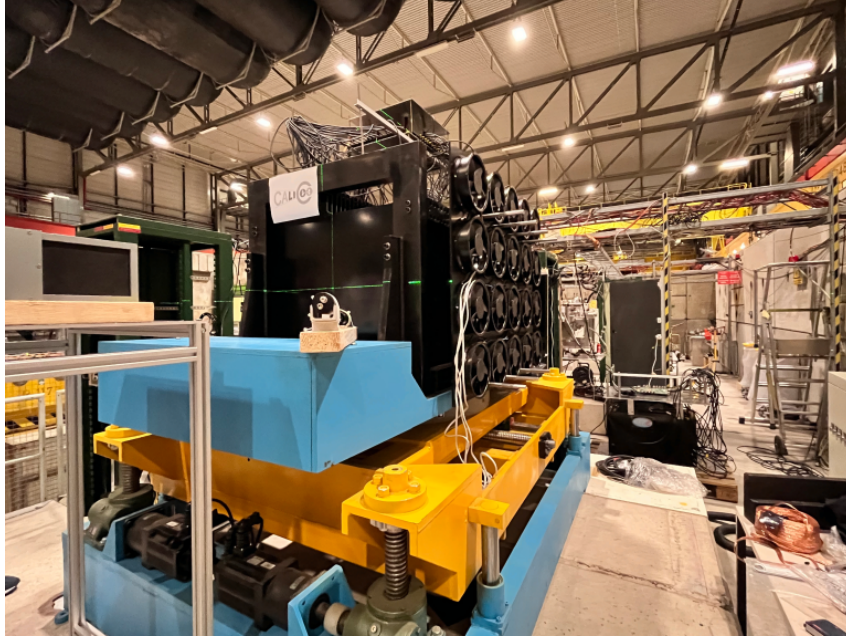


Figure 2: Photograph of the CEPC AHCAL prototype installed in the CERN beam-test area.

The experimental samples used in this study consist of positive-pion data collected at the H8 beam line. The H8 beam line provides positive-pion beams at different nominal momenta. Nine momentum points, corresponding to 10, 20, 30, 40, 50, 60, 70, 80, and 100 GeV/ $c$ , are selected for this analysis. For consistency with the Monte Carlo samples and the subsequent energy reconstruction results, the corresponding nominal energies are used to label the beam points throughout this work. A large number of raw events is available at each energy point, providing statistically independent samples for evaluating the machine learning models [1].

## 2.3 Monte Carlo Simulation and Detector Digitisation

Geant4 is used to simulate the transport and shower development of positive pions in the AHCAL [1]. The detector geometry in the simulation is constructed according to the beam-test prototype and contains the same 40 sampling layers composed of iron absorbers and plastic-scintillator active layers. The active area, scintillator-tile dimensions, and transverse segmentation of each layer are identical to those of the experimental prototype,

ensuring that the Monte Carlo samples and beam-test data share the same spatial readout structure.

Positive pions are generated using a particle gun. The particles are injected from a position in front of the first AHCAL layer and travel perpendicular to the active layers. Their transverse incident positions are sampled from a Gaussian distribution with a width parameter of 5 mm to approximate the beam-spot distribution on the prototype surface. For the training and validation samples, the incident energies are generated continuously between 1 and 120 GeV, enabling the models to learn the continuous relationship among visible energy, shower morphology, and true incident energy. In addition to these continuous-energy samples, statistically independent samples are generated at several fixed energies to evaluate the energy dependence of the model performance.

The energy deposits produced directly by Geant4 do not include the complete effects of scintillation-light production, SiPM response, and electronics readout and therefore cannot be compared directly with experimental data. To obtain Monte Carlo samples that approximate the actual AHCAL readout, the ideal Geant4 energy deposits are processed through a detector-response digitisation procedure. The digitisation includes four main stages: scintillator response, photoelectron production, SiPM response, and front-end electronics readout. The current digitisation model incorporates the principal detector-response effects, including scintillation photon statistics, scintillator response non-uniformity, SiPM photon detection efficiency, SiPM nonlinearity, gain fluctuations, electronic noise, and the readout threshold [1].

After digitisation, the Monte Carlo samples contain the energy and timing responses of individual AHCAL cells and can therefore be processed in the same manner as the beam-test data to construct the inputs to the machine learning models. Although the digitisation procedure includes the principal detector-response effects, the Monte Carlo simulation remains an approximation of the actual detector response [6]. Differences between simulation and beam-test data may arise from the hadronic interaction model, channel calibration uncertainties, electronics nonlinearity, noise, and experimental operating conditions [8]. The models are therefore evaluated separately using Monte Carlo and beam-test samples to investigate how these differences affect the performance of machine-learning-based energy reconstruction.

## 2.4 Dataset Construction

The datasets used in this study comprise continuous-energy Monte Carlo samples, single-energy Monte Carlo samples, and positive-pion beam-test samples. The continuous-energy samples are used to train and validate the machine learning models. The single-energy Monte Carlo samples are used to study the energy resolution and response linearity at different incident energies, while the beam-test samples are used to assess model generalisation to experimental detector data. The three types of samples are statistically independent, and neither the single-energy Monte Carlo samples nor the beam-test data are used to optimise the model parameters.

In the continuous-energy Monte Carlo samples, the incident energies of the positive pions are continuously distributed between 1 and 120 GeV. After preprocessing and event selection, the events are randomly divided into training and validation sets with a ratio of 4:1. The splitting is performed at the event level, and no event is shared between the two subsets. The training set is used to optimise the model parameters, while the validation set is used to monitor convergence, apply early stopping, and determine the optimal model checkpoint.

Independent single-energy Monte Carlo test samples are generated at nine fixed energies: 10, 20, 30, 40, 50, 60, 70, 80, and 100 GeV. Approximately 200,000 events are initially generated at each energy point. These samples are produced independently of the continuous-energy training and validation samples and are therefore statistically independent. At each fixed energy, the reconstructed-energy distribution predicted by a model is used to extract the peak position and distribution width, from which the energy response, energy resolution, relative improvement in resolution, and response linearity are determined.

The beam-test samples correspond to the same nine energy points as the single-energy Monte Carlo samples. The MLP and CNN models are evaluated using beam-test data at all nine energies. Because the beam-test data do not currently have a reliable timing calibration, no ParticleNet result is available for the beam-test samples. After channel calibration, threshold processing, and event selection, the events at each energy point are passed to the trained models. The model parameters are not updated during testing, and the beam-energy labels are not used for fine-tuning. The beam-test results therefore directly reflect the ability of the models to generalise from Monte Carlo simulations to

experimental data.

Data preprocessing is performed to improve the efficiency and stability of model training. First, Poisson fluctuations are introduced into the energy deposition of each active cell to provide a more realistic description of the experimental response. According to the minimum-ionising-particle (MIP) calibration of the AHCAL prototype, one MIP corresponds to a most probable energy deposition of approximately 0.461 MeV. The response threshold of each active cell is therefore set to 0.25 MeV, corresponding to approximately 0.5 MIP [1]. A cell with a visible energy less than or equal to this threshold is set to zero and treated as an unhit cell, whereas the energy, position, and timing information of a cell above the threshold is retained. This procedure suppresses the effects of electronic noise and low-energy random signals on the total visible energy and shower topology and ensures that a common hit definition is used for both the Monte Carlo and beam-test samples. After threshold processing, each event is converted into layer-wise statistical features, a regular three-dimensional energy tensor, or a hit-level point cloud according to the input requirements of the different models, as described in Section 3.

To reduce the effects of non-target particles, anomalous events, and severe energy leakage on model training and performance evaluation, common event-quality selection criteria are applied to the Monte Carlo and beam-test samples. The event selection consists of the following requirements.

1. Particle identification based on the fractal dimension. The beam-test samples may contain electron and muon contamination. The shower fractal dimension (FD) is therefore used for particle identification and is defined as

$$FD = \frac{\log R_\alpha}{\log \alpha} + 1, \quad R_\alpha = \frac{N}{N_\alpha} \quad (1)$$

where  $\alpha$  is the scale factor used to analyse the transverse shower pattern,  $N_\alpha$  is the number of hits at scale  $\alpha$ , and  $N$  is the total number of hits in the event.

Electron, muon, and hadron showers populate different regions of the FD distribution. Electron showers are relatively compact, muon events generally exhibit minimum-ionising-particle characteristics, and hadronic showers have more diffuse and irregular spatial structures. Linear selection boundaries are therefore applied in the corresponding two-dimensional feature space to reject electron-like and muon-like events while retaining positive-pion candidates in the hadronic region. Previous

studies have shown that this method can exploit the spatial shower patterns recorded by highly granular calorimeters to distinguish effectively among electrons, muons, and pions.

2. Selection based on the longitudinal shower position. The shower starting and ending layers are determined from the longitudinal distribution of the deposited energy. Events are required to have a shower starting layer below layer 10 and a shower ending layer below layer 35. Requiring the shower to start within the first ten layers rejects events in which the first inelastic interaction occurs deep inside the calorimeter and ensures that sufficient detector depth remains for shower development. Requiring the shower to end before layer 35 preserves a longitudinal margin at the rear of the AHCAL and reduces the impact of energy leakage beyond the detector. These requirements remove events with late shower initiation or incomplete longitudinal containment and thereby reduce the energy underestimation and reconstruction fluctuations associated with severe leakage.

The samples obtained after threshold processing and event selection are used for the subsequent model training and evaluation. All preprocessing parameters are determined exclusively from the training sample and are then fixed when applied to the validation, single-energy Monte Carlo test, and beam-test samples. This procedure prevents information from the test samples from entering the model training process.

### 3 Positive-Pion Energy Reconstruction Methods

This section presents the conventional and machine-learning-based methods used to reconstruct the energy of positive pions in the AHCAL. The conventional energy-sum method with global recalibration is first introduced and used as the baseline for subsequent performance comparisons. The energy reconstruction methods based on the MLP, CNN, and ParticleNet are then described, including their input representations and network architectures. Finally, the metrics used to evaluate the performance of the different reconstruction methods are defined.

### 3.1 Conventional Energy-Sum Method and Global Recalibration

When a charged particle passes through an AHCAL scintillator tile, it deposits energy that is converted by the SiPM into a measurable electrical signal. After electronics and MIP calibration, the response of each hit cell is expressed in MeV. For a given event, the most direct energy reconstruction method is to sum the calibrated energy deposits of all cells passing the readout threshold. The total deposited energy in the AHCAL is given by

$$E_{\text{HCAL}} = \sum_{i=1}^{N_{\text{hit}}} E_i \quad (2)$$

where  $N_{\text{hit}}$  is the number of cells with an energy deposition above the 0.5 MIP threshold, and  $E_i$  is the calibrated energy deposited in the  $i$ -th hit cell. Applying a hit threshold suppresses the effects of electronic noise and low-energy spurious signals on the reconstructed energy.

Because the AHCAL is a sampling calorimeter with a sampling fraction of only approximately 4%, a conversion between the total deposited energy and the incident particle energy is required. In this work, a global recalibration method is adopted. A common energy conversion factor  $\alpha$  is determined from the Monte Carlo samples and used to convert the total visible energy into the reconstructed energy:

$$E_{\text{rec}} = \alpha E_{\text{HCAL}} \quad (3)$$

The global recalibration factor is determined by minimising the overall difference between the reconstructed and true incident energies. Once obtained,  $\alpha$  is fixed and applied to all test samples.

The conventional energy-sum method is simple, physically transparent, and independent of complex model training. However, hadronic showers contain both electromagnetic and non-electromagnetic components and are affected by nuclear breakup, invisible energy losses, secondary-particle escape, and the non-compensating response of the detector. Two events with similar total visible energies may therefore have substantially different spatial distributions and shower compositions. A single global conversion factor cannot exploit event-level information such as the longitudinal shower profile, transverse spread, local energy density, and hit timing. It is consequently unable to correct hadronic shower fluctuations on an event-by-event basis [8]. In this study, the globally recalibrated energy-sum result is used as the baseline for evaluating the improvement achieved by the machine

learning models.

### 3.2 Energy Reconstruction Based on a Multilayer Perceptron

A multilayer perceptron (MLP) is one of the fundamental deep neural network architectures. It consists of an input layer, several fully connected hidden layers, and an output layer arranged sequentially. Each neuron in a given layer is connected to all neurons in the preceding layer, and nonlinear mappings are constructed through a combination of linear transformations and nonlinear activation functions [12]. In this study, the MLP is used as one of the machine learning approaches for hadron energy regression. Its input is a global feature vector extracted from the AHCAL response, and the model is trained to learn the mapping from calorimeter response features to the incident particle energy [12].

To obtain a fixed-dimensional input for the MLP, three physical quantities are extracted from each of the 40 AHCAL sampling layers: the total deposited energy  $E_{\text{Layer}}$ , the number of hit cells  $N_{\text{hit}}$ , and the average deposited energy density  $\rho = E/N$ . The total deposited energy in the full AHCAL,  $E_{\text{HCAL}}$ , is included as an additional global feature. Each event is therefore represented by a 121-dimensional feature vector:

$$\mathbf{x}_{\text{MLP}} = (E_1, \dots, E_{40}, N_1, \dots, N_{40}, \rho_1, \dots, \rho_{40}, E_{\text{HCAL}}) \quad (4)$$

The layer-wise energy deposits describe the longitudinal development of the hadronic shower, while the number of hits and the average deposited energy provide complementary information for distinguishing diffuse hadronic components from localised high-density energy deposits. Compared with the total visible energy alone, these features retain more information related to event-by-event shower fluctuations.

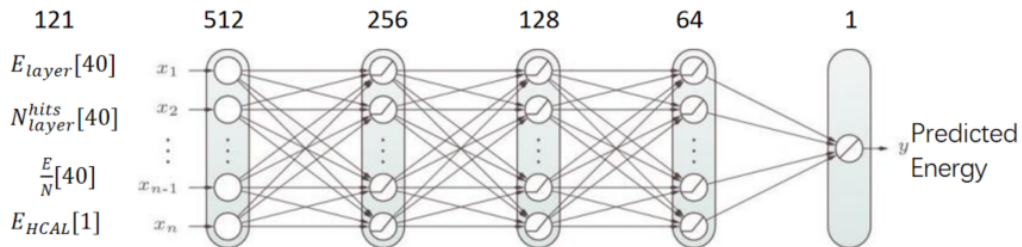


Figure 3: Schematic architecture of the MLP energy regression model.

The MLP contains four fully connected hidden layers. The 121-dimensional input vector is successively processed by hidden layers containing 512, 256, 128, and 64 neurons.

The output layer is a linear transformation that maps the final 64-dimensional feature vector to a single scalar corresponding to the reconstructed positive-pion energy.

The absolute scale of the reconstruction error increases with the incident energy. If the conventional mean squared error were used directly, high-energy events could dominate the total loss. To balance the contributions from different energy ranges during training, the MLP uses the mean squared relative error (MSRE) as its loss function:

$$\mathcal{L} = \frac{1}{N} \sum_{n=1}^N \frac{(\hat{E}_n - E_{\text{true},n})^2}{E_{\text{true},n}^2 + \epsilon} \quad (5)$$

where  $\hat{E}_n$  is the predicted energy of the  $n$ -th event,  $E_{\text{true},n}$  is the corresponding true incident energy, and  $\epsilon$  is a small constant for numerical stability. By constraining the relative rather than absolute reconstruction error, this loss gives comparable importance to low- and high-energy events and is therefore suitable for regression using mixed-energy training samples.

### 3.3 Energy Reconstruction Based on a Convolutional Neural Network

Although layer-wise statistical features describe the longitudinal development of a shower, they discard the transverse spatial distribution within each sampling layer. To exploit the highly granular AHCAL readout more fully, each event is represented as a three-dimensional calorimeter response tensor:

$$\mathbf{X} \in \mathbb{R}^{40 \times 18 \times 18} \quad (6)$$

where 40 is the number of longitudinal sampling layers and  $18 \times 18$  is the two-dimensional array of readout cells in each layer. Each element of the tensor corresponds to the energy deposited in a cell after threshold selection. This representation preserves both the transverse morphology and longitudinal evolution of the shower.

Hadronic showers exhibit complex structures in both the transverse and longitudinal directions. To exploit both types of information, an energy regression model combining a two-dimensional convolutional neural network and a bidirectional long short-term memory network is constructed [13]. The model uses the spatial feature extraction capability of the CNN to learn hadronic shower patterns directly from the three-dimensional AHCAL response. Convolutional operations capture local spatial correlations effectively, making

CNNs well suited to the rich energy-deposition patterns observed across the transverse and longitudinal calorimeter grids [13].

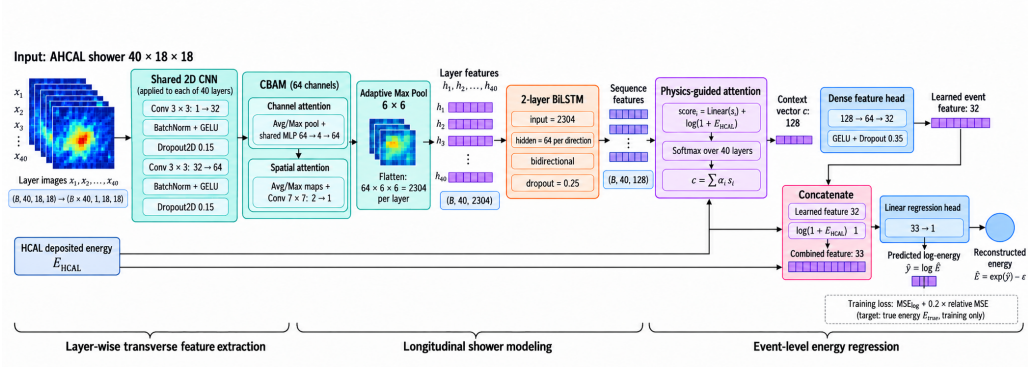


Figure 4: Schematic architecture of the CNN energy regression model.

As illustrated in figure 4, the CNN architecture consists of four main components: a two-dimensional convolutional feature extractor, a convolutional block attention module (CBAM), a long short-term memory (LSTM) sequence encoder, and a physics-guided attention module.

During convolutional feature extraction, the input tensor is divided along the layer dimension into a sequence of two-dimensional energy maps, which are processed individually by shared convolutional layers. The resulting feature maps are passed through the CBAM, which sequentially applies channel and spatial attention. This mechanism enhances the feature channels and spatial regions that contribute most strongly to energy regression while suppressing noise and less informative responses. The CBAM-enhanced feature maps are then compressed using adaptive pooling, reducing the computational cost of the subsequent modules while preserving the principal features of the transverse shower structure. The pooled features are arranged according to the sampling-layer sequence and passed to a two-layer bidirectional LSTM. LSTM networks are designed to capture long-range dependencies in sequential data. The bidirectional structure uses correlations in both longitudinal directions and thereby models the initiation, development, and attenuation of positive-pion showers in the AHCAL. Finally, the physics-guided attention module incorporates the total deposited energy as prior physical information when aggregating the longitudinal features. An attention score is first calculated for each longitudinal position, after which the total deposited energy is added to the score through a residual connection. The scores are normalised with a Softmax function and used to calculate a weighted sum of the layer features. The resulting aggregated feature vector is

transformed by fully connected layers, concatenated with the total deposited energy, and passed to the output layer to obtain the final energy prediction  $\hat{E}$ .

The CNN is trained using a combined loss function:

$$L_{\text{CNN}} = L_{\text{log}} + \lambda L_{\text{rel}} \quad (7)$$

where  $L_{\text{log}}$  is the mean squared error between the predicted and true energies in logarithmic space, and  $L_{\text{rel}}$  is the squared relative error in linear energy space. The logarithmic-space loss reduces the scale difference among different incident energies, while the relative-error term directly constrains the reconstruction bias in linear energy space.

By combining convolutional operations, attention mechanisms, and longitudinal sequence modelling, this architecture exploits the three-dimensional spatial information recorded by the AHCAL and learns both the transverse and longitudinal development of hadronic showers. Its comparison with the MLP and ParticleNet provides a means of studying how different levels of spatial information affect energy regression performance.

### 3.4 Energy Reconstruction Based on ParticleNet

The CNN represents the calorimeter response as a regular three-dimensional tensor that contains a large number of empty cells without valid signals. ParticleNet instead represents each shower as a point cloud composed only of effective hits, allowing the irregular spatial structure of a hadronic shower to be described directly at the hit level. Its central concept is to regard the detector hits as an unordered point cloud and use dynamic graph convolution to learn the geometrical relationships and local structural features among neighbouring hits [8,13]. This architecture is naturally suited to highly granular calorimeter data. Unlike conventional CNNs, which require the response to be mapped onto a regular grid, ParticleNet operates directly on the original hit cloud, avoiding information loss introduced by grid-based transformations and preserving the complete geometrical topology of the shower.

In this study, each hadronic shower is represented as an unordered set of hits. The  $i$ -th hit is described by a five-dimensional feature vector  $(x_i, y_i, z_i, E_i, t_i)$ , where  $x_i$ ,  $y_i$ , and  $z_i$  are the three-dimensional spatial coordinates,  $E_i$  is the deposited energy, and  $t_i$  is the hit time. Because the number of hits varies from event to event, the point clouds are padded or truncated to a fixed length during data loading. A mask is constructed to

identify valid hits and exclude padded points from the model calculations.

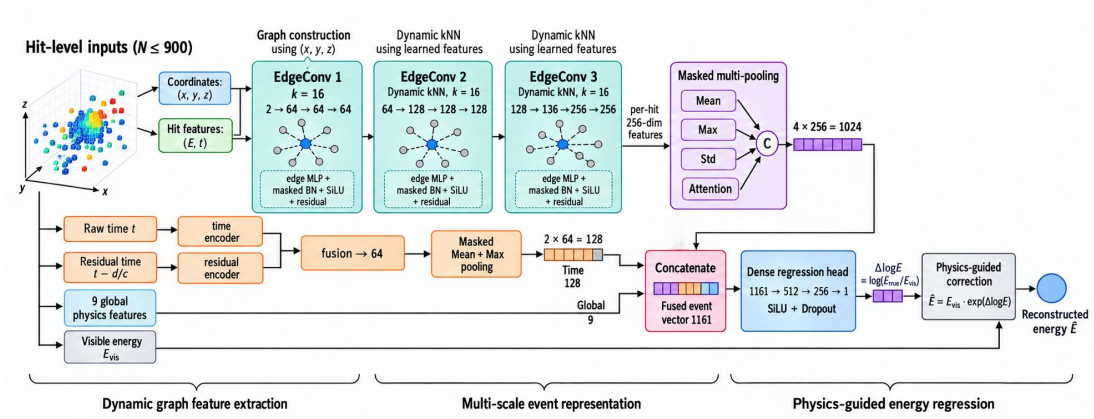


Figure 5: Schematic architecture of the ParticleNet energy regression model.

The backbone of ParticleNet consists of three stacked EdgeConv blocks. Within each block, a  $k$ -nearest-neighbour search is first performed using the spatial coordinates of the hits to construct a local graph. For each central hit, the model uses both its own features and the feature differences between the central hit and its neighbours to describe the local spatial and energy-deposition patterns. A shared convolutional network then aggregates the feature differences within each neighbourhood. By applying EdgeConv iteratively, the model learns progressively higher-level shower features and produces a fixed-dimensional global representation [7, 8].

An independent timing encoder is introduced to exploit the hit-time information. In addition to the original hit time, a time residual is calculated using the hit position and particle propagation distance to reduce the effect of differences in the time of flight [7]. The encoded timing features are fused with the point-cloud features extracted by the EdgeConv blocks. The spatial and timing features are then concatenated to form a fixed-dimensional representation of the complete event. This representation is processed by several fully connected layers with nonlinear transformations, and the final linear layer maps it to a single scalar corresponding to the predicted incident energy  $\hat{E}$ . The network can therefore be regarded as an end-to-end mapping,  $F_{\text{PN}} : \{(x_i, y_i, z_i, E_i, t_i)\}_{i=1}^N \rightarrow \mathbb{R}$ , which directly regresses the incident particle energy from the hit-level calorimeter response [7].

ParticleNet is trained using a combined loss consisting of a logarithmic correction loss and a relative energy loss:

$$L_{\text{PN}} = 0.7L_{\text{corr}} + 0.3L_{\text{energy}} \quad (8)$$

where  $L_{\text{corr}}$  constrains the difference between the predicted and true logarithmic correction factors, and  $L_{\text{energy}}$  constrains the relative deviation of the final reconstructed energy. Smooth L1 loss and Huber loss are used for the two terms, respectively, reducing the influence of a small number of anomalous shower events during training.

By representing showers as point clouds and combining dynamic graph convolution with timing information, ParticleNet is designed to exploit the high-granularity AHCAL information as fully as possible. Its comparison with the MLP and CNN reveals how different representations of shower geometry affect energy regression performance.

### 3.5 Performance Metrics

Three principal metrics are used to evaluate the models for hadron energy reconstruction: the energy resolution, the relative improvement over the baseline method, and the response linearity. All evaluations are performed using test samples grouped according to their true incident energies, allowing the energy dependence of the model performance to be studied.

At each incident energy, the reconstructed-energy distribution is fitted with a Crystal Ball function. The Crystal Ball function combines a Gaussian core with a low-energy power-law tail and provides an effective description of asymmetric calorimeter response distributions caused by energy leakage, invisible energy losses, and other non-Gaussian effects. Based on the fit results, the energy resolution is defined using the fitted mean and the full width at half maximum (FWHM):

$$\text{Resolution} = \frac{\text{FWHM}}{2.354 \times \text{Mean}} \times 100\% \quad (9)$$

Compared with resolution estimators that are more sensitive to the full distribution, the FWHM-based definition is less affected by the tails and provides a robust measure of the width of the central response peak. It is also consistent with the FWHM-based characterisation commonly used in experimental analyses, facilitating direct comparison with related measurements and simulations.

To quantify the performance gain of each machine learning model relative to the conventional method, the relative improvement is defined as

$$\text{Improvement} = \frac{\text{Res}_{\text{.sum}} - \text{Res}_{\text{.Model}}}{\text{Res}_{\text{.sum}}} \times 100\% \quad (10)$$

where  $\text{Res}_{\text{.sum}}$  is the energy resolution of the globally recalibrated energy-sum baseline defined in Section 3, and  $\text{Res}_{\text{.Model}}$  is the energy resolution of the MLP, CNN, or Parti-

cleNet at the same incident energy. A positive value indicates that the machine learning model improves upon the baseline. For example, an improvement of 50% means that the model reduces the energy resolution to one-half of the baseline value.

The response linearity is characterised by the relative difference between the central reconstructed energy and the true incident energy:

$$\frac{\Delta E}{E} = \frac{\text{Mean} - E_{\text{Truth}}}{E_{\text{Truth}}} \times 100\% \quad (11)$$

where Mean is the Gaussian mean obtained from the Crystal Ball fit and  $E_{\text{Truth}}$  is the corresponding true incident energy. This quantity indicates whether the model exhibits a systematic response bias. A value of zero corresponds to an ideal linear response, while positive and negative values indicate over-response and under-response, respectively. Response linearity is important for calorimeter energy calibration because an energy-scale bias directly affects the precision of jet energy reconstruction and boson mass measurements.

By comparing the reconstructed-energy distributions, energy resolutions, relative improvements, and response linearities, the different reconstruction methods can be evaluated comprehensively in terms of their response scale, distribution width, and consistency across the investigated energy range.

## 4 Energy Reconstruction Results

This chapter investigates the positive-pion energy reconstruction performance of the conventional energy-sum method and the three machine learning models. First, independent monoenergetic Monte Carlo samples are used to compare the reconstructed energy distributions, energy resolutions, relative improvements in resolution, and response linearities of the MLP, CNN, and ParticleNet models. The MLP and CNN models trained exclusively on Monte Carlo samples are then directly applied to CERN beam-test data to evaluate their performance under realistic detector conditions and to examine the differences between simulation and beam-test data. In all comparisons, the globally recalibrated conventional energy-sum method is taken as the baseline. This allows the ability of each model to correct event-by-event fluctuations in hadronic showers and to generalise from simulation to beam-test data to be evaluated comprehensively.

## 4.1 Reconstruction Results on Monte Carlo Samples

To evaluate the performance of the different energy reconstruction methods, independent monoenergetic positive-pion Monte Carlo samples were used for testing. The incident energies were 10, 20, 30, 40, 50, 60, 70, 80, and 100 GeV. At each energy point, the pion energy was reconstructed using the globally recalibrated conventional energy-sum method and the MLP, CNN, and ParticleNet models. Their performance was compared in terms of the reconstructed energy distribution, energy resolution, and response linearity. In the figures, the conventional energy-sum result is denoted by MC.

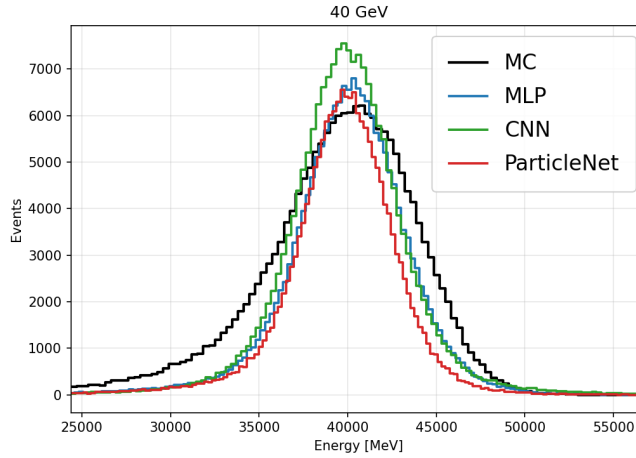


Figure 6: Reconstructed energy distributions for 40 GeV positive pions in the Monte Carlo sample.

Figure 6 shows the reconstructed energy distributions for the 40 GeV positive-pion sample. The distribution obtained using the conventional energy-sum method is relatively broad and exhibits a pronounced low-energy tail, reflecting the effects of invisible energy losses, energy leakage, and event-by-event fluctuations in hadronic showers [6, 7]. By contrast, the reconstructed energies obtained using all three machine learning methods are concentrated around 40 GeV, with substantially narrower distributions than that of the conventional method. This result indicates that the machine learning models can exploit information on the longitudinal development, transverse topology, and local energy deposition of the shower to correct fluctuations in the detector response on an event-by-event basis. Among the three models, ParticleNet produces the narrowest reconstructed energy distribution, followed by CNN and MLP. By directly processing hit-level point-cloud information, ParticleNet preserves the local spatial structure and energy distribution

of the shower more completely and therefore provides more effective event-by-event energy corrections [7, 8].

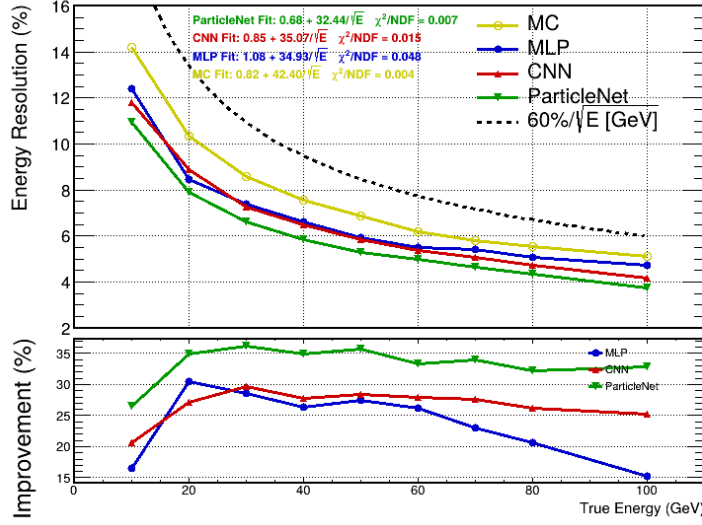


Figure 7: Energy resolutions and relative improvements of the different reconstruction methods on the Monte Carlo samples.

The upper panel of Figure 7 compares the energy resolutions of the four reconstruction methods at different incident energies. The energy resolution improves progressively with increasing incident energy for all methods. At 10 GeV, the energy resolutions obtained with the conventional method, MLP, CNN, and ParticleNet are 14.19%, 12.41%, 11.79%, and 10.93%, respectively. At 100 GeV, the corresponding values are 5.11%, 4.72%, 4.17%, and 3.74%. All three machine learning methods outperform the conventional energy-sum method at every tested energy, with ParticleNet consistently providing the best energy resolution. With the exception of the 20 GeV point, CNN also achieves a better energy resolution than MLP.

The dependence of the energy resolution on the incident energy is fitted using

$$R(E) = p_0 + \frac{p_1}{\sqrt{E}} \quad (12)$$

where  $E$  is expressed in GeV. The fit result for the conventional method is

$$R_{MC}(E) = 0.82 + \frac{42.40}{\sqrt{E}} \quad (13)$$

whereas those for MLP, CNN, and ParticleNet are, respectively,

$$R_{MLP}(E) = 1.08 + \frac{34.93}{\sqrt{E}} \quad (14)$$

$$R_{\text{CNN}}(E) = 0.85 + \frac{35.07}{\sqrt{E}} \quad (15)$$

$$R_{\text{PN}}(E) = 0.68 + \frac{32.44}{\sqrt{E}} \quad (16)$$

Compared with the conventional method, all three machine learning methods reduce the stochastic term, demonstrating that they can effectively suppress event-by-event fluctuations in hadronic showers. ParticleNet gives the smallest stochastic and constant terms among the four methods, indicating favourable energy reconstruction performance over the full energy range considered.

The lower panel of Figure 7 shows the relative improvements in energy resolution achieved by the three machine learning methods with respect to the conventional method. The improvement obtained with MLP ranges from 15.19% to 30.45%, with an average improvement of approximately 23.80% over the nine energy points. CNN yields improvements ranging from 20.65% to 29.73%, with an average of approximately 26.74%. ParticleNet provides improvements ranging from 26.47% to 36.23%, with an average of approximately 33.39%. MLP achieves a marked improvement in the low- and intermediate-energy regions, although its relative improvement decreases at higher energies. The improvement provided by CNN remains relatively stable over the full energy range. ParticleNet achieves the largest improvement at every tested energy and reaches a maximum improvement of approximately 36.23% at 30 GeV.

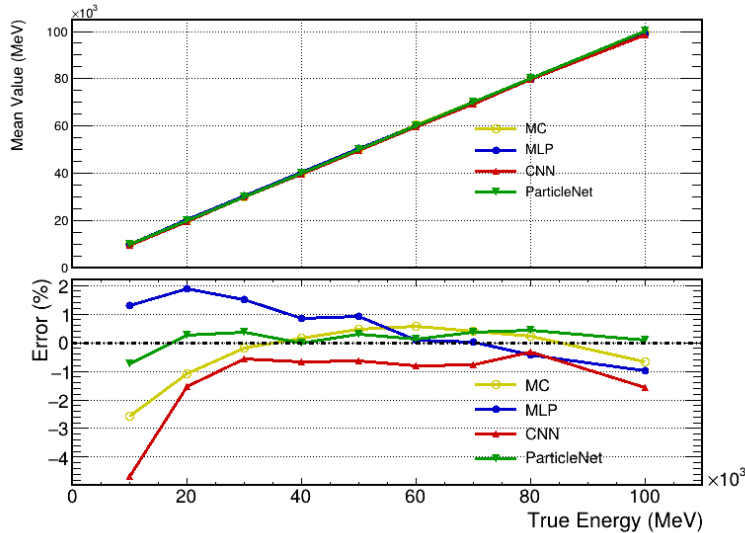


Figure 8: Mean reconstructed energy and relative response deviation as functions of the true incident energy for the Monte Carlo samples.

Figure 8 shows the mean reconstructed energy as a function of the true incident energy, together with the corresponding relative response deviation. Overall, the mean reconstructed energies obtained with all four methods vary approximately linearly with the incident energy, indicating that both the conventional method and the machine learning models correctly reproduce the overall relationship between the calorimeter response and the incident energy. Nevertheless, differences among the methods remain, particularly in the low- and high-energy regions.

For the conventional energy-sum method, the relative response deviation ranges from  $-2.57\%$  to  $0.57\%$ . The incident energy is underestimated by approximately  $2.57\%$  and  $1.08\%$  at 10 and 20 GeV, respectively. For MLP, the relative deviation ranges from  $-0.98\%$  to  $1.88\%$ . The model slightly overestimates the incident energy between 10 and 50 GeV, after which the response gradually changes to a small underestimation as the energy increases. The CNN predictions are generally lower than the true energies, with relative deviations ranging from  $-4.66\%$  to  $-0.31\%$ . The largest deviation occurs at 10 GeV; above 10 GeV, the deviation remains within  $-1.59\%$ . These results show that, although CNN substantially improves the energy resolution, a systematic response bias remains in the low-energy region.

ParticleNet exhibits the best response linearity among the four methods. Its relative deviation remains between  $-0.73\%$  and  $0.44\%$  over the entire tested energy range, with a mean absolute deviation of approximately  $0.30\%$ . Except at 10 GeV, the deviation is below  $0.45\%$  at all energy points. This result indicates that predicting an event-by-event correction with respect to the total visible energy can improve the energy resolution while preserving good response linearity.

In summary, all three machine learning models improve the positive-pion energy reconstruction performance on the Monte Carlo samples. By using layer-wise statistical features, MLP already provides a clear improvement over the conventional method. CNN further exploits the spatial structure of the calorimeter response and improves the energy resolution at most energy points. By modelling the hadronic shower using hit-level point clouds, dynamic graphs, and global physical features, ParticleNet achieves the best energy resolution and response linearity at all tested energies. It provides an average resolution improvement of approximately  $33.39\%$  and a mean absolute response deviation of approximately  $0.30\%$ , demonstrating its strong capability for positive-pion energy

reconstruction.

## 4.2 Reconstruction Results on Beam-Test Data

To examine the applicability of the machine learning models to real detector data, the models trained on continuous-energy Monte Carlo samples were directly applied to the CERN beam-test data. The test samples comprise positive-pion beams at nine energies: 10, 20, 30, 40, 50, 60, 70, 80, and 100 GeV. The same hit threshold and event-selection criteria were applied to both the beam-test and Monte Carlo samples, and the globally re-calibrated conventional energy-sum result was used as the performance baseline. Because reliable timing calibration is not available for the beam-test data, ParticleNet results on these data are not currently available. This section therefore focuses on the beam-test performance of MLP and CNN.

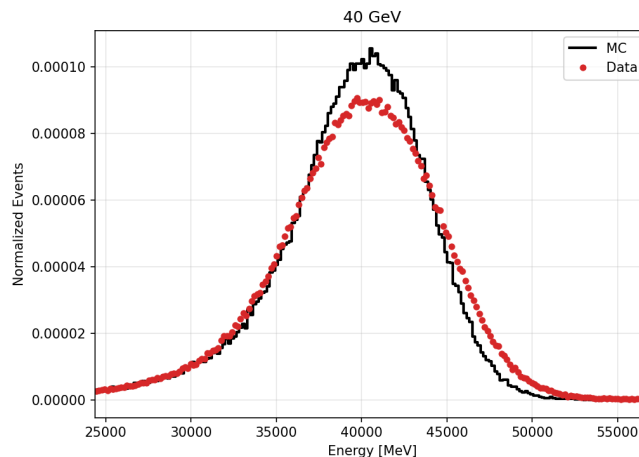


Figure 9: Comparison of the conventional reconstructed energy distributions for the 40 GeV Monte Carlo and beam-test samples.

Figure 9 compares the conventional reconstructed energy distributions of the Monte Carlo and beam-test samples at 40 GeV. The distribution centres of both samples are located near 40 GeV, indicating that the detector digitisation procedure provides a reasonable description of the average energy response observed in the beam test. Nevertheless, the beam-test distribution is broader than the Monte Carlo distribution and exhibits a more pronounced high-energy tail. The differences between beam-test data and simulation may arise from channel-to-channel response nonuniformities, electronic noise, SiPM non-linearity, the beam-energy spread, and limitations of the detector digitisation model [6, 8].

These effects can produce discrepancies between the simulated and measured distributions and make the direct application of models trained exclusively on Monte Carlo samples more challenging.

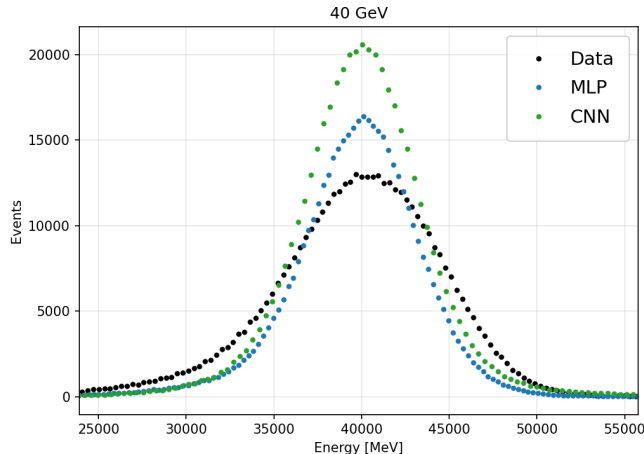


Figure 10: Reconstructed energy distributions obtained with the conventional method, MLP, and CNN for the 40 GeV beam-test sample.

Figure 10 further shows the reconstructed energy distributions obtained with MLP and CNN for the 40 GeV beam-test sample. The energy resolution obtained using the conventional energy-sum method is 10.80%, whereas those obtained with MLP and CNN are reduced to 7.66% and 7.64%, respectively. These values correspond to relative improvements of 29.10% and 29.29%. The distribution centres produced by both machine learning models remain close to 40 GeV, while the distribution widths and low-energy tails are substantially reduced. This observation demonstrates that the shower features learned from Monte Carlo samples remain effective for beam-test data and can correct event-by-event response fluctuations in real detector measurements.

The upper panel of Figure 11 presents the energy resolutions of the conventional method, MLP, and CNN at different beam energies. The energy resolution improves progressively with increasing incident energy for all three methods. At 10 GeV, the energy resolution is 17.48% for the conventional method, 13.62% for MLP, and 13.53% for CNN. At 100 GeV, the corresponding values are 5.84%, 4.68%, and 4.86%. MLP and CNN both outperform the conventional method at all nine beam-energy points.

The energy resolutions are fitted using

$$R(E) = p_0 + \frac{p_1}{\sqrt{E}} \quad (17)$$

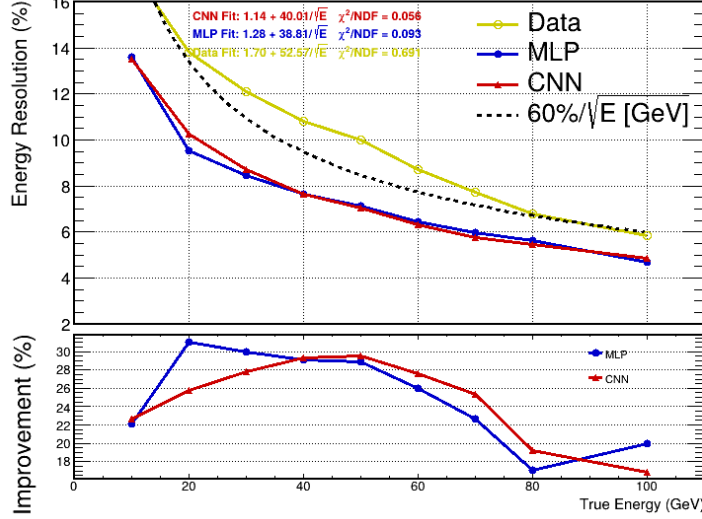


Figure 11: Energy resolutions and relative improvements of the different reconstruction methods on the beam-test data.

where  $E$  is expressed in GeV. The fit result for the conventional method is

$$R_{\text{Data}}(E) = 1.70 + \frac{52.67}{\sqrt{E}} \quad (18)$$

whereas those for MLP and CNN are, respectively,

$$R_{\text{MLP}}(E) = 1.28 + \frac{38.81}{\sqrt{E}} \quad (19)$$

$$R_{\text{CNN}}(E) = 1.14 + \frac{40.01}{\sqrt{E}} \quad (20)$$

Compared with the conventional method, both machine learning models substantially reduce the stochastic term, indicating that they can exploit information on the shower topology and longitudinal development to reduce the impact of hadronic-shower fluctuations on energy reconstruction. MLP gives a slightly smaller stochastic term, whereas CNN has a relatively smaller constant term. Overall, the two models exhibit comparable performance over the full beam-energy range.

The lower panel of Figure 11 shows the relative improvement in energy resolution achieved by the two models with respect to the conventional method. The improvement obtained with MLP ranges from 17.10% to 31.08%, with an average of approximately 25.19% over the nine energy points. For CNN, the improvement ranges from 16.81% to 29.53%, with an average of approximately 24.88%. MLP provides better energy resolutions at 20, 30, and 100 GeV, whereas CNN slightly outperforms MLP at 10 GeV and at most

energy points between 40 and 80 GeV. The overall difference between the two models is small, and both improve the energy resolution of the beam-test data by approximately 25% on average.

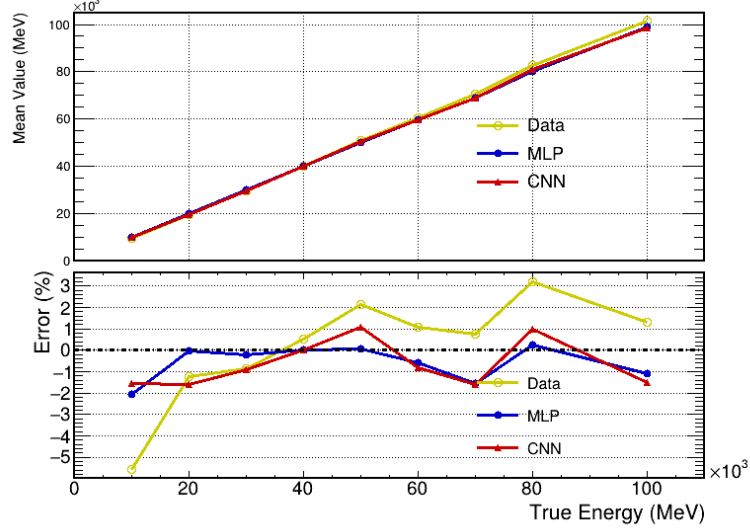


Figure 12: Mean reconstructed energy and relative response deviation as functions of the beam energy.

Figure 12 shows the mean reconstructed energy as a function of the beam energy, together with the corresponding relative response deviation. For the conventional method, the relative deviation ranges from  $-5.55\%$  to  $3.19\%$ . At 10 GeV, the reconstructed energy is underestimated by approximately 5.55%, whereas a certain degree of overestimation is observed at some energy points of 50 GeV and above. The mean absolute deviation is approximately 1.85%, indicating that a noticeable energy dependence remains after global recalibration.

For MLP, the relative response deviation ranges from  $-2.05\%$  to  $0.24\%$ , with a mean absolute deviation of approximately 0.66%. Except at 10 GeV, the deviation remains within approximately 1.54% and is close to zero between 20 and 50 GeV. For CNN, the relative deviation ranges from  $-1.62\%$  to  $1.07\%$ , with a mean absolute deviation of approximately 1.11%. CNN exhibits a smaller deviation than MLP at 10 GeV and an almost vanishing response bias at 40 GeV, although systematic variations of approximately 1% remain at several intermediate- and high-energy points. Overall, both machine learning models improve the response linearity of the beam-test data, with MLP giving the smaller mean absolute deviation over the full energy range.

Compared with the monoenergetic Monte Carlo test results, the energy resolutions of both models are slightly degraded on the beam-test data. On average, the MLP and CNN resolutions on the beam-test data are approximately 0.84 and 1.11 percentage points higher, respectively, than those obtained on the Monte Carlo samples. For MLP, the average performance difference between the beam-test data and Monte Carlo simulation is approximately 0.83%, consistent with the expected discrepancy between the simulated and actual detector responses. Nevertheless, without retraining on beam-test data, both models still provide substantial improvements over the conventional method at every beam-energy point while maintaining good response linearity. These results demonstrate that the models possess a certain degree of simulation-to-data transferability.

In summary, both MLP and CNN improve the positive-pion energy resolution on the beam-test data by approximately 25% on average while keeping the response deviations within approximately 2% at most energy points. Their overall resolution performances are comparable: CNN performs slightly better at several intermediate- and high-energy points, whereas MLP exhibits a smaller mean response bias. The performance on the beam-test data is slightly poorer than that obtained in the Monte Carlo simulation, indicating that discrepancies remain between the detector digitisation model and the actual detector response. Future improvements may be achieved by refining the digitisation procedure, further optimising event selection, and introducing domain-adaptation techniques between simulated and beam-test data.

## 5 Summary and Outlook

High-granularity calorimeters provide detailed measurements of the spatial and temporal structure of hadronic showers, offering rich information for machine-learning-based energy reconstruction. This study focuses on pion events in the CEPC AHCAL prototype and addresses the limitation of the conventional energy-sum method in correcting hadronic-shower fluctuations. Three machine learning models—MLP, CNN, and ParticleNet—were developed, and their performance was investigated using Monte Carlo simulation and CERN beam-test data.

Monte Carlo simulation and detector digitisation were performed for the AHCAL prototype. Continuous-energy samples were used for model training, while independent

monoenergetic samples were used for testing. Data preprocessing included noise suppression using a 0.5 MIP threshold, event selection based on the shower fractal dimension and hit energy, and constraints on the longitudinal shower range. The MLP model uses layer-wise statistical features and provides a simple and efficient architecture. The CNN model takes a three-dimensional tensor as input and combines convolutional operations, attention mechanisms, and a bidirectional LSTM to extract transverse and longitudinal shower features. ParticleNet represents valid hits as a point cloud and employs dynamic graphs and EdgeConv to learn local spatial relationships. The three models therefore exploit the high-granularity calorimeter information from complementary perspectives.

Results obtained with the Monte Carlo samples show that all three models significantly outperform the conventional method in energy resolution, with average relative improvements of 23.80%, 26.74%, and 33.39% for MLP, CNN, and ParticleNet, respectively. ParticleNet achieves the best performance, improving the energy resolution by 26.47%–36.23% over the full energy range while maintaining good response linearity, with a mean absolute deviation of only 0.30%. CNN generally performs slightly better than MLP, although a small systematic bias remains in the low-energy region. In the beam-test study, the MLP and CNN models trained exclusively on simulated data were directly applied to experimental data. They achieved average improvements in energy resolution of 25.19% and 24.88%, respectively, with mean absolute response deviations of 0.66% and 1.11%. Despite small discrepancies between the simulated and experimental results, both models substantially improve the reconstruction performance on real detector data, demonstrating the feasibility of the proposed methods.

In summary, the spatial shower information recorded by a high-granularity calorimeter effectively supplements the event-level information that is not captured by the total visible energy. Machine learning models can learn correlations between the shower topology and invisible energy losses or energy leakage, thereby suppressing the effects of hadronic-shower fluctuations. The reconstruction performance improves as the input representation progresses from statistical quantities to three-dimensional images and subsequently to point clouds, with ParticleNet providing the best overall performance.

Future work will further improve pion energy reconstruction and refine the Monte Carlo simulation and detector digitisation procedures to reduce discrepancies between simulated and experimental data. Ensemble learning and Bayesian neural networks may

be investigated to provide event-level uncertainty estimates, while attention and feature analyses may help clarify the physical origin of the performance improvements and enhance model reliability. The application of these methods will also be extended to more complex physics events, providing technical support for particle-flow reconstruction, jet measurements, and calorimeter optimisation at the CEPC.

## References

- [1] Y. Shi. *Study and development of the CEPC analog hadronic calorimeter*. PhD thesis, University of Science and Technology of China, 2024.
- [2] Y. Dong, M. Ruan, K. Wang, H. Yang, and J. Zhu. Testing a 95 GeV scalar at the CEPC with machine learning. *Chin. Phys. C*, 50:031001, 2026.
- [3] CMS Collaboration. Full event interpretation with machine-learning-based particle-flow reconstruction in the CMS detector. *CERN-EP-2025-300*, 2026.
- [4] X. Xia, B. Qi, M. Ruan, Y. Liu, and Y. Che. Pattern recognition at CEPC AHCAL prototype using beam test data at CERN. *EPJ Web Conf.*, 320:00047, 2025.
- [5] X. Li, H. Diao, and Y. Zhang. Study on electromagnetic response and energy reconstruction of high-granularity calorimeter based on machine learning. Unpublished manuscript, 2026.
- [6] CMS HGCAL and CALICE AHCAL Collaborations. Using graph neural networks to reconstruct charged pion showers in the CMS high granularity calorimeter. *JINST*, 19:P11025, 2024.
- [7] T. Yang, Y. Wang, and M. Ruan. Machine-learning-based energy reconstruction for hadronic showers. Unpublished manuscript, 2026.
- [8] C.L. Liao et al. Application of machine learning methods for proton energy reconstruction on space based high granularity calorimeter. *Exp. Astron.*, 62:7, 2026.
- [9] H. Lu. Searching for supersymmetric dark matter using convolutional neural networks. Master’s thesis, Nanjing Normal University, 2022. In Chinese.
- [10] Y. Li, Y. Song, X.-R. Lyu, and Y. Zheng. Deep learning enables anti-neutron reconstruction in the EMC, 2025. Presentation at the BESIII P&S Workshop, 26 September 2025.
- [11] C. Li. *Research on beta-gamma particle-discrimination phoswich detector based on pulse-shape discrimination technology*. PhD thesis, Jilin University, 2023. In Chinese.

- [12] S. Schnake. Energy regression with deep learning in particle collider physics. Master's thesis, Universitat Hamburg, 2019.
- [13] Y. Shi and V. Boudry. Optimisation of a silicon-tungsten electromagnetic calorimeter energy response to photons. *JINST*, 21:P08028, 2026.