Overview of Machine Learning Applications at the Pierre Auger Observatory

Ezequiel Rodriguez[1,2] for the Pierre Auger Collaboration[3*]

[1]Instituto de Tecnologías en Detección y Astropartículas, San Martín, Argentina
[2]Karlsruher Institut für Technologie, Karlsruhe, Germany
[3]Observatorio Pierre Auger, Av. San Martín Norte 304, Malargüe, Argentina
Full author list: http://www.auger.org/archive/authors_2024_09.html

Abstract. The complex spatio-temporal information from shower footprints, comprised of particle arrival times and traces measured by water-Cherenkov detectors, is challenging to analyse with traditional methods but well-suited for machine learning (ML) based analyses. In this contribution, we provide an overview of the ML applications developed to leverage the high event statistics acquired by the Pierre Auger Observatory. In the context of the energy spectrum, a neural network approach for energy reconstruction has demonstrated potential in reducing composition biases in the energy estimator. A notable application for mass composition is the indirect prediction of the depth of the maximum shower development, $X_{\max}$, which extends the energy range of previous analyses into unexplored higher energies. Aligned with AugerPrime, the ongoing upgrade of the Observatory, the impact of enhanced electronics and scintillation detectors was explored via simulations. Both transformers and convolutional networks perform better at the reconstruction of mass-composition sensitive observables like $X_{\max}$ and the muon number, demonstrating the benefits of the Observatory's upgrade.

1 Introduction

Since 2004, the Pierre Auger Observatory is the largest detector for extensive air showers induced by ultra-high-energy cosmic rays. Machine learning (ML) offers new tools to enhance research in physical sciences. This work highlights three applications to Observatory data.

2 Data

The Observatory is a hybrid air-shower detector comprised of a Surface Detector (SD) and a Fluorescence Detector (FD) [1]. The main component of the SD is the SD-1500, which is a triangular grid of over 1600 water-Cherenkov detectors (WCDs) spaced 1500 meters apart ($\sim 100\%$ duty cycle). Photomultiplier tubes in the WCDs measure Cherenkov light of the particles in the shower front traversing the water volume. The FD consists of 27 fluorescence telescopes across four sites, and directly observe the longitudinal profile and energy of the shower cascade during clear, moonless nights (15% duty cycle). Neural Networks (NNs) have been trained to leverage the complex spatio-temporal data from SD shower footprints. The training data sets for all presented NNs consist of detector simulations of air showers induced by proton, helium, oxygen, and iron nuclei. Refer to Refs [2–6] for details.

*e-mail: spokespersons@auger.org

3 Machine Learning Applications

3.1 Energy Reconstruction

The standard energy reconstruction used for SD data neglects the temporal evolution from the signals measured at the ground. Hence, a NN that exploits the time signals from the WCDs was developed and tested. Descriptions of the input, training procedure, and architecture are available in [2]. The NN outperforms the standard energy estimation in simulations, reducing the primary mass-dependent bias from ±10% to ±4%, and improving the resolution by 5% at the highest energies. Systematic discrepancies between observed data and simulations required the correction of detector aging effects in the time signals and calibration of the network's output with hybrid SD-FD events. After removing remaining biases, the NN reconstruction was found comparable to the standard approach.

3.2 Indirect measurement of X_{max}

The mass composition of cosmic rays was studied using a NN estimator of the depth of the shower maximum X_{max}. The indirect measurement of X_{max} with the SD provides much higher event statistics than direct FD measurements, extending the energy range of previous analyses into unexplored higher energies. Details of the input, training, and architecture can be found in [4]. After calibration using high-quality SD-FD events, the network predictions were benchmarked against simulations of pure proton and iron compositions. The results display a clear trend with energy, showing a transition from lighter (proton-like) to heavier (iron-like) compositions as energy increases, being in agreement with direct measurements of the first and second moment of the FD [5].

3.3 Machine Learning in the AugerPrime era

The Observatory is undergoing an upgrade (AugerPrime) [7]. Key improvements include the installation of a surface scintillation detector (SSD) on top of each WCD, and the change to new electronics that triple the digitization rate. To assess the impact of the new detector system, mass-sensitive observables were estimated using NNs. Transformers and CNNs were trained to predict X_{max} and the muon number R_{μ}. Details about the inputs, training, and architecture are available in Refs [3, 6]. In the case of the X_{max} estimator, the transformers demonstrate improved resolution with simulations incorporating the upgraded detectors. When adding SSD information, both NN models exhibit an improved resolution for R_{μ} predictions, particularly for more vertical showers, due to the SSD's larger effective area.

4 Summary

The Pierre Auger Collaboration has successfully implemented ML methods focusing on the estimation of high-level shower observables. Studies with NNs suggest that the new SSDs of AugerPrime will enhance the mass separation capabilities of the Observatory.

References

[1] Aab A. *et al.*† 2015 *Nucl. Instrum. Meth. A* **798** 172–213
[2] Ellwanger F. *et al.*† *PoS*(ICRC2023) **275**
[3] Hahn S. T. *et al.*† *PoS*(ICRC2023) **318**
[4] Abdul Halim A. *et al.*† *Phys. Rev. Lett.* (in press) 2024 arXiv:2406.06319
[5] Abdul Halim A. *et al.*† *Phys. Rev. D* (in press) 2024 arXiv:2406.06315
[6] Langner N. *et al.*† PoS(ICRC2023) **371**
[7] Aab A. *et al.*† 2016 arXiv:1604.03637
†: (Pierre Auger Collaboration)