The Study of Intelligent Algorithm in Particle Identification of Heavy-Ion Collisions at Low and Intermediate Energies

Authors: Gao-Yi Cheng, Qian-Min Su, Xi-Guang Cao, Guo-Qiang Zhang, Qian-Min Su, Xi-Guang Cao

Date: 2024-01-10T00:00:00+00:00

Abstract

Traditional particle identification methods are time consuming, experience-dependent, and poor repeatability challenges in heavy-ion collisions at low and intermediate energies. Researchers urgently need solutions to the dilemma of traditional particle identification methods. This study explores the possibility of applying intelligent learning algorithms to the particle identification of heavy-ion collisions at low and intermediate energies. Multiple intelligence algorithms, including XgBoost and TabNet, were selected to test datasets from the neutron ion multi-detector for reaction-oriented dynamics (NIMROD–ISiS) and Geant4 simulation. Machine learning algorithms based on tree structures and deep learning algorithms e.g. TabNet show excellent performance and generalization ability. Adding additional data features besides energy deposition can improve the algorithm's identification ability when the data distribution is nonuniform. Intelligent learning algorithms can be applied to solve the particle identification problem in heavy-ion collisions at low and intermediate energies.

Full Text

The Study of Intelligent Algorithms for Particle Identification in Heavy-Ion Collisions at Low and Intermediate Energies

Gao-Yi Cheng,[1,2,3] **Qian-Min Su,**[1,†] **Xi-Guang Cao,**[2,3,‡] **and Guo-Qiang Zhang**[2,3]

[1]School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

[2]Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China
[3]Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201800, China

Traditional particle identification methods in heavy-ion collisions at low and intermediate energies suffer from being time-consuming, experience-dependent, and having poor repeatability. Researchers urgently need solutions to overcome these limitations. This study explores the application of intelligent learning algorithms to particle identification in heavy-ion collisions at low and intermediate energies. Multiple algorithms, including XgBoost and TabNet, were tested on datasets from the Neutron Ion Multi-detector for Reaction-Oriented Dynamics (NIMROD–ISiS) and Geant4 simulations. Machine learning algorithms based on tree structures and deep learning algorithms such as TabNet demonstrated excellent performance and generalization capability. Adding supplementary data features beyond energy deposition improved algorithm identification accuracy when dealing with non-uniform data distributions. These results demonstrate that intelligent learning algorithms can effectively solve particle identification problems in heavy-ion collisions at low and intermediate energies.

Keywords: Heavy-ion collisions at low and intermediate energies, Machine learning, Ensemble learning algorithm, Particle identification, Data imbalance

INTRODUCTION

Intelligent algorithms play crucial roles in nuclear physics. Nuclear physics experiments face challenges including high complexity, extensive data volumes, time-consuming operations, and intricate models. For example, particle collision experiments generate millions of terabytes of data daily in high-energy heavy-ion collisions. Consequently, extracting useful information from complex experimental data has become an enormous challenge.

Large-scale experiments such as ATLAS, ALICE, and CMS have already applied machine learning and deep learning algorithms [?] to analyze and process experimental data. Typical examples include research on particle track reconstruction problems [?] in high-energy physics experiments, as well as data analysis and pattern recognition for the Higgs boson [?]. Comprehensive reviews of machine learning applications in particle physics are available through the HEPML Living Review [?] and the ML Physics Portal website [?].

Currently, research on intelligent algorithms in nuclear physics experiments [?] focuses on data analysis for nuclear masses [?], nuclear charge radii [?], decay half-lives [?], critical reaction thresholds [?], and spallation reaction cross-sections [?]. In addition to using machine learning algorithms to investigate various physical issues [?], researchers have applied these algorithms to analyze experimental data directly [?], involving tasks such as particle trajectory reconstruction, vertex reconstruction [?], and particle identification in nuclear reactions. Advancements in experimental equipment and related technologies have facilitated the integration of machine learning and nuclear physics.

Current research on particle identification primarily focuses on high-energy particle physics. To date, studies have mainly concentrated on identifying particle types [?] and separating rare particles from background signals. The data and algorithms used for particle identification depend on the detector type. For example, the output data from calorimeter detectors can be processed and converted into matrix data, enabling the use of image algorithms such as CNN and GNN for processing.

Research and applications of machine learning in particle identification have primarily focused on LHC detectors, such as calorimeters [?] and Cherenkov detectors [?]. Moreover, a recent research focus in LHC experiments has been the development of new detector software and hardware based on machine learning and deep learning algorithms [?].

Compared with other nuclear reactions, the particles generated in heavy-ion collisions at low and intermediate energies are of various types and have complex energy distributions. Numerous fragments have similar charges and masses. Experiments on heavy-ion collisions depend on the energy resolution of the detector and require a detection array with large solid-angle coverage. Therefore, identifying dozens or even hundreds of reaction products from independent detection units is challenging. Traditional particle identification methods include telescope [?], time-of-flight [?], magnetic spectrometer, Bragg spectroscopy, and pulse shape discrimination methods. These methods are often combined to improve identification capability, especially for heavy fragments with minor differences in charge and mass numbers between adjacent fragments. The performance of traditional methods for heavier particles is hindered by their dependence on experience, poor repeatability, and time consumption. Precise identification of charge and mass numbers is fundamental to all research related to heavy-ion collisions and provides a powerful method for studying exotic nuclear configurations [?]. Compared with particle identification in high-energy physics, the wide variety and slight differences in the charge and mass numbers of charged particles produced in heavy-ion reactions pose significant challenges for existing particle identification methods. Therefore, developing a universal, efficient, and high-precision particle identification method based on machine learning techniques will significantly boost the study of heavy-ion collisions.

Parker et al. [?] devised a 5-layer neural network and evaluated its performance on the 22nd and 23rd detectors of NIMROD–ISiS. We also utilized a dataset from the NIMROD–ISiS detector array. This study aimed to identify particle charge and mass numbers in heavy-ion collisions at low and intermediate energies. Supervised learning algorithms were used to train particle identification models based on ΔE-E energy deposits from telescope (or super-telescope) detectors in heavy-ion collisions. Machine learning and deep learning algorithms were applied to identify particle charge and mass numbers, and their capabilities were compared.

II. DATASET AND METHODS

2.1 Data Sources

Real-world data (RWD) were obtained from experiments on heavy-ion collisions at low and intermediate energies carried out at the Cyclotron Institute of Texas A&M University, consisting of reaction products detected by the NIMROD–ISiS array [?, ?]. In addition to the dataset from Texas A&M University, Geant4 [?] was used to simulate heavy-ion collisions at intermediate energies. The QMD model with G4IonQMDPhysics was employed as an event generator to simulate the reaction process of a beam incident on a target. The detection processes in Geant4 include electromagnetic interactions (G4EmStandardPhysics), energy transfer and loss (G4EmExtraPhysics and G4StoppingPhysics), decay processes (G4DecayPhysics and G4RadioactiveDecayPhysics), and elastic and inelastic scattering (G4HadronHElasticPhysics, G4HadronPhysicsINCLXX, and G4IonElasticPhysics). The simulations involved collisions of ^{28}Si with an energy of 50 MeV/u and ^{12}C particles in vacuum. The detector system consisted of four super-telescope detectors, generating a dataset with more than four million particles.

2.2 Detector Description

The NIMROD–ISiS detector array comprised 14 rings with experimental data obtained from 143 detectors, including 124 telescope detectors and 19 super-telescope detectors with ring numbers ranging from 2 to 15. The detection system included Si detectors and CsI(Tl) scintillators covering angles from 3.6° to 167.0°. The back half of NIMROD (90.0°–167.0°) consists of half the Indiana Silicon Sphere. Si detectors were combined with CsI detectors as "telescopes," while some were equipped with two Si detectors in tandem, known as "super-telescopes" (3.6°–45°), enhancing the ability to identify mass numbers of heavier fragments. The capacity to include ionization chambers in front of Si detectors is also available. [Figure 1: see original paper] shows the structure of the NIMROD-ISiS detector array.

[Figure 1: see original paper] Schematic diagram of the NIMROD–ISiS detector array layout (from Texas A&M NIMROD–ISiS official website).

[Figure 2: see original paper] shows the structure of the super-telescope detector used in the Geant4 simulation and the ΔE-E two-dimensional histogram from Geant4 simulation.

2.3 Machine Learning Algorithms

This study covers several common machine learning algorithms, including Support Vector Machines (SVM), Logistic Regression (LR), and Bayesian classifiers. Ensemble learning algorithms based on tree structures and TabNet, a deep learning algorithm, were also employed. The algorithms used in this study are briefly described below.

(a) MLP
Multi-layer perceptron (MLP) is a feed-forward neural network composed of multiple neurons, forming the basis and prototype of many artificial and deep learning neural networks.

(b) Random Forest
Random forest is an early tree-based ensemble learning algorithm with multiple decision trees [?], offering advantages of both decision trees and ensemble learning, including strong robustness and predictive capability.

(c) XgBoost
XgBoost is a tree-based ensemble learning algorithm proposed in 2016 [?] that is widely used in data mining, natural language processing, image recognition, and other fields. Generally, XgBoost is a machine learning algorithm with high efficiency, accuracy, flexibility, explainability, and scalability.

(d) LightGBM
LightGBM, a tree-based gradient boosting framework for ensemble learning, has been widely used in various applications [?]. Built on the gradient-boosting decision tree (GBDT) algorithm, LightGBM incorporates advanced techniques such as gradient-based one-sided sampling (GOSS) and histogram-based acceleration. These optimizations enable faster training and lower memory consumption, making LightGBM an efficient and practical choice for machine learning tasks.

(e) CatBoost
CatBoost is a tree-based ensemble learning algorithm developed by Yandex [?]. Compared with XgBoost and LightGBM, CatBoost can automatically process categorical features and handle feature scaling without additional data preprocessing. CatBoost adopts the same gradient-based splitting and feature selection strategies based on a greedy algorithm as XgBoost. CatBoost also automatically handles missing values without additional data padding and demonstrates robustness to noise and outliers.

Boosting-based ensemble learning algorithms such as XgBoost, LightGBM, and CatBoost are widely used in various fields. The basic process involves training multiple weak learners, assigning weights to training samples, and iteratively adjusting these weights based on learner performance to create a powerful ensemble model capable of accurate classification. [Figure 3: see original paper] depicts the underlying structure of ensemble learning algorithms employing the boosting method.

(f) TabNet
TabNet, introduced by Google in 2019, is a neural network architecture explicitly designed for classification, prediction, and regression tasks involving tabular data [?]. Unlike traditional decision-tree-based machine learning algorithms, TabNet minimizes the need for preprocessing input data and can automatically learn interdependencies among input features. It incorporates an attention transformer that uses an attention mechanism to dynamically select

relevant feature vectors. Since its inception, TabNet has been widely adopted in various applications involving tabular data [?, ?].

2.4 Methodology and Training Process

[Figure 4: see original paper] illustrates the procedure for applying intelligent algorithms in this study. Training a classification model typically involves several steps: (a) Data acquisition: obtaining a dataset containing particle charge and mass information from experimental or simulated data; (b) Data preprocessing: ensuring data quality and consistency through noise removal, addressing missing data, and normalizing features; (c) Data splitting: dividing the dataset into training and testing sets, with random and stratified sampling being commonly used methods; (d) Feature engineering: transforming, extracting, and selecting raw data to create informative feature sets; (e) Algorithm selection: choosing suitable algorithms based on specific task requirements, with the main task being multi-classification; (f) Training and parameter tuning: adjusting algorithm parameters to improve model performance, with each algorithm having a unique set of adjustable parameters; and (g) Performance assessment: evaluating the trained model using the testing dataset with appropriate evaluation metrics for particle identification.

Based on the NIMROD–ISiS structure, the dataset was initially split according to the ring number determined by the detector's forward angle. Subsequently, data were divided into two categories: telescope and super-telescope detectors. The Geant4 dataset was used for training and testing with machine learning and deep learning algorithms. After identifying optimal algorithms, a subset of detector data was used to evaluate algorithm generalization capability.

Two classification strategies were adopted: (a) training and testing charge and mass numbers independently, and (b) including the particle's charge number as part of the data features when classifying mass numbers. From a logical perspective, the second strategy resembles traditional particle identification methods.

In practical studies, experimental data exhibited highly unbalanced distributions. Random extraction can lead to disparate category distributions among training, validation, and test sets, causing critical errors and poor performance. Therefore, stratified sampling was employed as an alternative to random sampling to address this problem.

III. RESULTS AND DISCUSSION

3.1 Evaluation Metrics

As the core task of particle identification involves multi-class classification, using suitable evaluation metrics is crucial. Common metrics include accuracy, recall, precision, and F1-score [?], which assess algorithm performance from different aspects. Classification results can be categorized into four types: (a) true positive (TP): predicting positive samples as positive; (b) true negative (TN):

predicting negative samples as negative; (c) false positive (FP): predicting negative samples as positive; and (d) false negative (FN): predicting positive samples as negative.

When evaluating algorithms, corresponding metrics are calculated using classification results through the following equations:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Accuracy measures overall correctness, precision measures the accuracy of positive predictions, recall represents the coverage of correctly predicted positive samples, and the F1-score is a compound metric combining precision and recall.

Since multi-classification tasks extend the binary positive-negative problem to multiple categories, comprehensive evaluation methods are essential. Commonly used strategies include micro-averaging, macro-averaging, and weighted averaging. Macro-averaging calculates the average precision and recall for each class. Micro-averaging ignores category differences and calculates overall TP, FP, TN, and FN values. Weighted averaging is similar to macro-averaging but uses category proportions as weights.

In particle identification, all generated particles have equal significance. Therefore, macro-averaging was selected as the calculation method for evaluation metrics, providing a balanced assessment across all classes and facilitating comprehensive understanding of model performance. Since particle categories were determined by charge and mass numbers, these were merged into a binary data format to calculate evaluation metrics.

3.2 Results on NIMROD–ISiS Data

The particles detected by the NIMROD–ISiS detector array can be categorized as light ions (with proton numbers ranging from one to four) or heavy ions. Most heavy ions cannot penetrate the Si detector, whereas most light particles pass through it. The disparity in production yield between light particles and heavy ions during the reaction process leads to data distribution imbalances. The dataset was split based on whether particles hit the CsI detector, solving the data imbalance problem and improving algorithm performance.

The XgBoost ensemble learning algorithm was selected for testing, with input features including total energy, energy deposition in Si and CsI detectors, and detector position. Particle charge and mass numbers served as data labels. [Figure 5: see original paper] shows model results based on telescope data, while presents results on super-telescope data using two different classification strategies.

[Figure 5: see original paper] Test results of XgBoost on NIMROD–ISiS telescope data. Figures (a) and (b) show results for particles without registration on CsI detectors, with the latter performing better than the former.

The model performed well when tested on particles not registered on CsI detectors, with evaluation metrics for each ring generally exceeding 0.85. The model achieved higher accuracy in identifying charge numbers than mass numbers. While also achieving high accuracy for particles registered on CsI detectors, their precision, recall, and F1-score were low. This discrepancy is attributed to extreme data imbalance. [Figure 6: see original paper] shows the mass distribution for ring 2, revealing that the mass distribution of particles registered on CsI detectors was highly non-uniform, with significant differences between the most and least frequent categories. Precise identification of rare categories remains challenging for this model. Since evaluation uses macro-averaging, classifier metrics are calculated as average values across all categories, meaning performance on low-percentage categories significantly affects overall metrics.

The training strategies for charge and mass numbers showed no significant differences. Including the charge number as an additional feature did not effectively improve identification ability for particles with deficient quantities. If the model fails to precisely predict charge numbers, mass number identification accuracy is also affected.

[Figure 6: see original paper] Mass number distribution for events without registration on CsI detectors in ring 2. The distribution of particles without CsI registration is well balanced, with sample sizes exceeding 1000 in most categories. Among particles registered on CsI detectors, most heavy ions have only around 100 occurrences.

3.3 Addressing Data Imbalance

To address this problem, several methods were proposed: (a) algorithm parameter optimization—refining parameters (reducing learning rate, increasing iterations, expanding tree depth) to improve accuracy, precision, and recall, though parameter adjustment alone had limited impact on low-sample categories even with distinct weights; (b) data category adjustment—reducing the imbalance ratio by eliminating categories with only a few or few dozen samples; and (c) exploration of data preprocessing methods—trying different approaches including normalization, standardization, or no preprocessing.

The most effective solution for severe sample shortages in specific categories is including additional data, which reduces the imbalance ratio and improves accuracy. For instance, in ring 10, each category had over 20,000 samples with an imbalance ratio of only 5:1, and XgBoost performed excellently with evaluation metrics exceeding 0.9 for all types.

Other factors such as detector position and hardware conditions (temperature and electronic signal drift) can cause scaling issues, affecting algorithm accuracy. To address this, Geant4 was used to simulate experiments and detector performance, enabling focused research on the imbalance issue.

3.4 Geant4 Simulation Results

In Geant4, total particle energy, time of flight (ToF), kinetic energy before entering the detector, detector position, and particle deposition energy (Eabs) were selected as input features. Testing with XgBoost demonstrated that additional features alleviated the data imbalance problem, resulting in excellent performance.

To confirm that these results were not limited to XgBoost alone, comparative tests were conducted with other machine learning and deep learning algorithms. The results ([Figure 7: see original paper]) confirmed earlier findings: tree-based algorithms such as XgBoost and deep learning TabNet demonstrated excellent performance, whereas traditional algorithms such as LR, SVM, and Bayesian classifiers exhibited poor performance. These results validate the effectiveness of the proposed approach in mitigating data imbalances and highlight the superiority of tree-based and deep learning algorithms in addressing this challenge.

[Figure 7: see original paper] Test results of machine learning and deep learning algorithms on the Geant4 dataset. SVM, MNB, GNB, and LR perform poorly, while MLP has relatively low precision and recall. Ensemble learning algorithms such as XgBoost and deep learning algorithm TabNet perform well.

3.5 Feature Engineering Analysis

Subsequently, algorithms were evaluated using only energy deposition as a feature. XgBoost, LightGBM, CatBoost, and TabNet, which exhibited promising performance in prior tests, were selected for this assessment. Results demonstrated decreased accuracy in predicting particle mass numbers when using only this feature.

Based on these observations, a series of additional features were selected for comparative analysis. Numerous tests showed that particle flight time is important for improving algorithm accuracy. Although this feature alone proved insufficient for charge and mass number identification, its combination with energy deposition enhanced identification capabilities. Detailed test results are presented in and .

Classification results from Geant4 simulation data, with independent training on charge and mass numbers.

Classification results from Geant4 simulation data, where charge number is included as a data feature for mass number classification.

3.6 Generalization Ability Tests

The final phase involved comprehensive investigation of algorithm generalization capability. Unlike previous tests using data from all detectors, this phase focused on training models using specific detector subsets while reserving remaining data for testing. Features included time-of-flight (ToF) and energy deposition. Various preprocessing techniques, including normalization and standardization, were explored. Before training, datasets from specific detectors were normalized and standardized using MinMaxScaler and StandardScaler methods from sklearn.preprocessing, with these methods also applied to test data from other detectors before evaluation. However, these methods significantly reduced generalization capability, so no data preprocessing was performed. Results are shown in [Figure 8: see original paper].

Algorithm performance was excellent, with TabNet and XgBoost evaluation metrics mostly exceeding 0.9 for all detector data ([Figure 9: see original paper]). These findings establish the efficacy of training models with robust generalization abilities even with limited data availability, highlighting the advantages and effectiveness of machine learning and deep learning algorithms and demonstrating their potential for practical applications.

[Figure 8: see original paper] Generalization ability test results of the XgBoost algorithm on the Geant4 dataset using different data preprocessing methods. Both normalization and standardization severely reduce model generalization ability.

[Figure 9: see original paper] Generalization ability test results for XgBoost, CatBoost, LightGBM, and TabNet. Figures (a), (b), (c), and (d) show accuracy, precision, recall, and F1-score curves, respectively. All algorithm evaluation metrics exceed 0.8, with XgBoost and TabNet mostly exceeding 0.9. TabNet shows better generalization ability than ensemble learning algorithms.

Inspired by these findings, a similar study of data similarity was conducted on specific NIMROD–ISiS rings. Input features included total energy, energy deposition, and detector position. Testing revealed that detectors could be grouped based on data similarity. Taking data from ring 9 (particles registered on CsI detectors) as an example, ring 9 can be divided into two detector groups ([Figure 10: see original paper]a and [Figure 10: see original paper]b). These results provide valuable information on NIMROD–ISiS detector patterns and characteristics, contributing to algorithm optimization and data processing method improvements. These findings also have significant implications for detector array design and performance enhancement.

[Figure 10: see original paper] Tests on NIMROD–ISiS ring 9 show that predictions for other detector data can be categorized into two scenarios, (a) and (b).

A high degree of similarity is observed between detectors on ring 9.

IV. CONCLUSION

Particle identification in machine learning is an integrated problem requiring consideration of various factors including data selection, partitioning, feature engineering, preprocessing, algorithm selection, and parameter tuning. Traditional methods require significant manual effort and are limited by researcher experience and available time. This study aimed to develop a universal and adaptable particle identification model to assist manual processes. While achieving 100% accuracy may not be possible, ensemble learning algorithms produced meaningful results, particularly XgBoost.

The conclusions are as follows:

First, intelligent algorithms, particularly tree-based ensemble learning algorithms, can effectively identify particles in heavy-ion collisions at low and intermediate energies, offering a viable alternative to traditional methods.

Second, addressing data imbalances is crucial for particle identification. Severe imbalances significantly affect results. Solutions include ensuring sufficient data for balanced distributions, adding supplementary features beyond particle energy deposition, and constructing different identification models based on detector structure.

Third, training specialized particle identification models using existing data reduces the time and resources required for traditional identification. This benefits laboratories conducting long-term, large-scale heavy-ion collision experiments and paves the way for developing professional particle identification software.

Finally, machine learning algorithms can be used to study detector similarity, particularly in large-scale detector arrays with complex structures.

Future studies should explore combinations of supervised and unsupervised learning approaches. Other physics software such as NpTool [?] will be used to simulate experiments, as it is known for efficient project management and simulation of sophisticated detector arrays.

Since Geant4 simulations are time-consuming and resource-intensive, alternative approaches for generating particle collision data are needed. Generative Adversarial Networks (GAN) [?] and Variational Autoencoders (VAE) [?] have shown promise in generating simulated detector data in high-energy physics [?]. Utilizing GAN and VAE can reduce the time and resources required for massive simulated data generation, making the process more efficient and accessible.

Building on TabNet's excellent performance, further investigations will include additional deep learning algorithms such as DeepGBM [?] and GrowNet [?]. Moreover, we will attempt to modify existing ensemble learning algorithms into multi-output algorithms to classify mass and charge numbers simultaneously. Our research aims to enhance understanding of detector systems in sophisticated experiments, which can be used to explore interesting clustering phenomena in nuclei [**?**].

V. ACKNOWLEDGEMENTS

The Cyclotron Institute of Texas A&M University is gratefully acknowledged for permission to use NIMROD–ISiS data to train and test machine learning algorithms. We are deeply indebted to Professor J. B. Natowitz (Texas A&M University) for valuable suggestions and critical review of this manuscript.

REFERENCES

[1] A. Kalweit, Particle identification in the ALICE experiment. J. Phys. G. Nucl. Partic. 38, 124073 (2011).
[2] C. Zampolli, Particle identification with the ALICE detector at the LHC. (2012).
[3] P. Križan, Particle identification at Belle II. J. INSTRUM. 9, C07018 (2014).
[4] Ł.K. Graczykowski, M. Jakubowska, K.R. Deja et al., Using machine learning for particle identification in ALICE. J. INSTRUM. 17, C07016 (2022).
[5] P. Calafiura, S. Farrell, H. Gray et al., TrackML: a high energy physics particle tracking challenge. IEEE 14th International Conference on E-Science (e-Science) 344-344 (2018).
[6] C. Tüysüz, F. Carminati, B. Demirköz et al., Particle track reconstruction with quantum algorithms. Epj. Web. Conf. 245, 09013 (2020).
[7] O. Bakina, D. Baranov, I. Denisenko et al., Deep learning for track recognition in pixel and strip-based particle detectors. J. INSTRUM. 17, P12023 (2022).
[8] P. Goncharov, E. Schavelev, A. Nikolskaya et al., Ariadne: PyTorch library for particle track reconstruction using deep learning. AIP Conf. Proc. 2377, (2021).
[9] T.Q. Chen, T. He, Higgs boson discovery with boosted trees. JMLR Work. Conf. Proc. 42, 69-80 (2015).
[10] M. Azhari, A. Abarda, B. Ettaki et al., Higgs boson discovery using machine learning methods with PySpark. Procedia Comput. Sci. 170, 1141-1146 (2020).
[11] C. Adam-Bourdarios, G. Cowan, C. Germain-Renaud et al., The Higgs machine learning challenge. J. Phys.: Conf. Ser. 664, 072015 (2015).
[12] S.R. Ahmad, Technical report of participation in Higgs boson machine learning challenge. (2015).
[13] A.E. Phoboo, Machine learning wins the Higgs challenge. (CERN Bulletin, 2014), https://cds.cern.ch/journal/CERNBulletin/2014/49/News%20Articles/1972036. Accessed November 16, 2023.
[14] University of California, Irvine. ML Physics Portal, http://mlphysics.ics.uci.edu/. Accessed November 16, 2023.
[15] M.J. Fenton, A. Shmakov, T.W. Ho et al., Permutationless many-jet event reconstruction with symmetry preserving attention networks. Phys. Rev. D

105, 112008 (2022).
[16] C. Shimmin, P. Sadowski, P. Baldi et al., Decorrelated jet substructure tagging using adversarial neural networks. Phys. Rev. D 96, 074034 (2017).
[17] P. Baldi, P. Sadowski, D. Whiteson, Searching for exotic particles in high-energy physics with deep learning. Nat. Commun. 5, 4308 (2014).
[18] W.B. He, J.J. He, R. Wang et al., Machine learning applications in nuclear physics (in Chinese). Sci. Sin.-Phys. Mech. Astron. 52, 252004 (2022).
[19] W.B. He, Q.F. Li, Y.G. Ma et al., Machine learning in nuclear physics at low and intermediate energies. Sci. China Phys. Mech. Astron. 66, 282001 (2023).
[20] M. Zhou, Y.Q. Luo, H.C. Song, Applications of machine learning in relativistic heavy ion physics (in Chinese). Sci. Sin.-Phys. Mech. Astron. 52, 252002 (2022).
[21] W.B. He, Y.G. Ma, L.G. Pang et al., High-energy nuclear physics meets machine learning. Nucl. Sci. Tech. 34, 88 (2023).
[22] T.L. Zhao, H.F. Zhang, Neural network approach to improve the quality of atomic nuclei (in Chinese). Sci. Sin.-Phys. Mech. Astron. 52, 252008 (2022).
[23] M.R. Mumpower, T.M. Sprouse, A.E. Lovell et al., Physically interpretable machine learning for nuclear masses. Phys. Rev. C 106, L021301 (2022).
[24] Z.M. Niu, H.Z. Liang, Nuclear mass predictions with machine learning reaching the accuracy required by r-process studies. Phys. Rev. C 106, L021303 (2022).
[25] L. Neufcourt, Y.C. Cao, W. Nazarewicz et al., Bayesian approach to model-based extrapolation of nuclear observables. Phys. Rev. C 98, 034318 (2018).
[26] Z.P. Gao, Y.J. Wang, H.L. Lü et al., Machine learning the nuclear mass. Nucl. Sci. Tech. 32, 109 (2021).
[27] J.Q. Ma, Z.H. Zhang, Improved phenomenological nuclear charge radius formulae with kernel ridge regression. Chinese Phys. C 46, 074105 (2022).
[28] X.X. Dong, R. An, J.X. Lu et al., Nuclear charge radii in Bayesian neural networks revisited. Phys. Lett. B 838, 137726 (2023).
[29] R. Utama, W.C. Chen, J. Piekarewicz, Nuclear charge radii: density functional theory meets Bayesian neural networks. J. Phys. G. Nucl. Partic. 43, 114002 (2016).
[30] S.J. Tao, L.F. Zhang, Q.Y. Zhang et al., Improved naive Bayesian probability classifier in nuclear charge radius prediction (in Chinese). Sci. Sin.-Phys. Mech. Astron. 52, 252009(2022).
[31] Y.F. Ma, C. Su, J. Liu et al., Predictions of nuclear charge radii and physical interpretations based on the naive Bayesian probability classifier. Phys. Rev. C 101, 014304 (2020).
[32] Z.M. Niu, H.Z. Liang, B.H. Sun et al., Predictions of nuclear β-decay half-lives with machine learning and their impact on γ-process nucleosynthesis. Phys. Rev. C 99, 064307 (2019).
[33] J.M. Munoz, S. Akkoyun, Z.P. Reyes et al., Predicting β-decay energy with machine learning. Phys. Rev. C 107, 034308 (2023).
[34] Z.Y. Yuan, D. Bai, Z.Z. Ren et al., Theoretical predictions on α-decay

properties of some unknown neutron-deficient actinide nuclei using machine learning. Chinese Phys. C 46, 024101 (2022).
[35] X.D. Bu, D. Wu, C.L. Bai, Prediction of α-decay half-lives for superheavy nuclei based on neural network (in Chinese). Sci. Sin.-Phys. Mech. Astron. 52, 252005 (2022).
[36] P. Li, J.H. Bai, Z.M. Niu et al., β-decay half-lives studied using neural network method (in Chinese). Sci. Sin.-Phys. Mech. Astron. 52, 252006 (2022).
[37] N.J. Costiris, E. Mavrommatis, K.A. Gernoth et al., Decoding β-decay systematics: a global statistical model for β-half-lives. Phys. Rev. C 80, 044332 (2009).
[38] R. Wang, Y.G. Ma, R. Wada et al., Nuclear liquid-gas phase transition with machine learning. Phys. Research 2, 043202 (2020).
[39] D. Peng, H.L. Wei, J. Pu et al., Bayesian neural network prediction methods for fragment cross sections in proton-induced spallation reactions (in Chinese). Sci. Sin.-Phys. Mech. Astron. 52, 252012 (2022).
[40] B.C. Wang, M.T. Qiu, W. Chen et al., Machine learning-based analyses for total ionizing dose effects in bipolar junction transistors. Nucl. Sci. Tech. 33, 131 (2022).
[41] Y.B. Yu, G.F. Liu, W. Xu et al., Research on tune feedback of the Hefei Light Source II based on machine learning. Nucl. Sci. Tech. 33, 28 (2022).
[42] Y.D. Song, R. Wang, Y.G. Ma et al., Determining the temperature in heavy-ion collisions with multiplicity distribution. Phys. Lett. B 814, 136084 (2021).
[43] Q.F. Song, L. Zhu, J. Su, Target dependence of isotopic cross sections in the spallation reactions 238U + p, d and 9Be at 1 AGeV. Chinese Phys. C 46, 074108 (2022).
[44] H.K. Wu, Y.J. Wang, Y.M. Wang et al., Machine learning method for 12C event classification and reconstruction in the active target time-projection chamber. Nucl. Instrum. Meth. A (2023).
[45] F.P. Li, Y.J. Wang, Z.P. Gao et al., Application of machine learning in the determination of impact parameter in the 132Sn + 124Sn system. Phys. Rev. C 104, 034608 (2021).
[46] Z.Y. Li, Z. Qian, J.H. He et al., Improvement of machine learning-based vertex reconstruction for large liquid scintillator detectors with multiple types of PMTs. Nucl. Sci. Tech. 33, 93 (2022).
[47] H. Arahmane, E.M. Hamzaoui, Y.B. Maissa et al., Neutron-gamma discrimination method based on blind source separation and machine learning. Nucl. Sci. Tech. 32, 18 (2021).
[48] J. Collado, J.N. Howard, T. Faucett et al., Learning to identify electrons. Phys. Rev. D 103, 116028 (2021).
[49] L. de Oliveira, B. Nachman, M. Paganini, Electromagnetic showers beyond shower shapes. Nucl. Instrum. Meth. A 951, 162879 (2020).
[50] P. Baldi, K. Bauer, C. Eng et al., Jet substructure classification in high-energy physics with deep neural networks. Phys. Rev. D 93, 094034 (2016).

[51] C. Fanelli, J. Pomponi, DeepRICH: learning deeply Cherenkov detectors. Mach. Learn.: Sci. Technol. 1, 015010 (2020).
[52] E. Cisbani, A. Del Dotto, C. Fanelli et al., AI-optimized detector design for the future Electron-Ion Collider: the dual-radiator RICH case. J. INSTRUM. 15, P05009 (2020).
[53] S. Carboni, S. Barlini, L. Bardelli et al., Particle identification using the ΔE–E technique and pulse shape discrimination with the silicon detectors of the FAZIA project. Nucl. Instrum. Meth. A 664, 251-263 (2012).
[54] W. Klempt, Review of particle identification by time of flight techniques. Nucl. Instrum. Meth. A 433, 542-553 (1999).
[55] Y.G. Ma, Effects of α-clustering structure on nuclear reaction and relativistic heavy-ion collisions. Nuclear Techniques 46, 080001 (2023).
[56] J.J. He, W.B. He, Y.G. Ma et al., Machine-learning-based identification for initial clustering structure in relativistic heavy-ion collisions. Phys. Rev. C 104, 044902 (2021).
[57] Y.G. Ma, S. Zhang, Influence of nuclear structure in relativistic heavy-ion collisions. Handbook of Nuclear Physics 1-30 (2022).
[58] X.G. Cao, E.J. Kim, K. Schmidt et al., Examination of evidence for resonances at high excitation energy in the 7α disassembly of 28Si. Phys. Rev. C 99, 014606 (2019).
[59] X.G. Cao, E.J. Kim, K. Schmidt et al., α and α conjugate fragment decay from the disassembly of 28Si at very high excitation energy. JPS Conf. Proc. 010038 (2020).
[60] X.G. Cao, E.J. Kim, K. Schmidt et al., Evidence for resonances in the 7α disassembly of 28Si. AIP Conf. Proc. 2038, 020021 (2018).
[61] P. Adamson, M. Youngs, Machine learning: potential application for particle identification. 2019 Fall Meeting of the APS Division of Nuclear Physics (2019).
[62] S. Wuenschel, K. Hagel, R. Wada et al., NIMROD–ISiS, a versatile tool for studying the isotopic degree of freedom in heavy ion collisions. Nucl. Instrum. Meth. A 604, 578-583 (2009).
[63] R. Wada, S. Wuenschel, K. Hagel et al., A 4π detector array, NIMROD-ISIS. Nucl. Phys. News 24, 28-33 (2014).
[64] S. Agostinelli, J. Allison, K. Amako et al., GEANT4—a simulation toolkit. Nucl. Instrum. Meth. A 506, 250-303 (2003).
[65] L. Breiman, Random forests. Mach. Learn. 45, 5-32 (2001).
[66] T.Q. Chen, C. Guestrin, XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining 785-794 (2016).
[67] G.L. Ke, Q. Meng, T. Finley et al., LightGBM: a highly efficient gradient boosting decision tree. Adv. Neur. In. 30, (2017).
[68] A.V. Dorogush, V. Ershov, A. Gulin, CatBoost: gradient boosting with categorical features support. (2018).
[69] S. Ö. Arik, T. Pfister, Tabnet: attentive interpretable tabular learning. In Proceedings of the AAAI Conference on Artificial Intelligence 35, 6679-6687 (2021).

[70] J.Z. Yan, T.Y. Xu, Y.C. Yu et al., Rainfall forecast model based on the tabnet model. Water 13, 1272 (2021).
[71] R. Asencios, C. Asencios, E. Ramos, Profit scoring for credit unions using the multilayer perceptron, XGBoost and TabNet algorithms: evidence from Peru. Expert Syst. Appl. 213, 119201 (2023).
[72] B. Juba, H.S. Le, Precision-recall versus accuracy and the role of large data sets. Proceedings of the AAAI Conference on Artificial Intelligence 33, 4039-4048 (2019).
[73] N. Japkowicz, Assessment metrics for imbalanced learning. Imbalanced Learning: Foundations, Algorithms, and Applications 187-206 (2013).
[74] M. Grandini, E. Bagli, G. Visani, Metrics for multi-class classification: an overview. (2020).
[75] M. Hossin, M.N. Sulaiman, A review on evaluation metrics for data classification evaluations. IJDKP. 5, 1 (2015).
[76] E. Mortaz, Imbalance accuracy metric for model selection in multi-class imbalance classification problems. Knowl.-Based Syst. 210, 106490 (2020).
[77] A, Matta, P. Morfouace, N. de Séréville et al., NPTool: a simulation and analysis framework for low-energy nuclear physics experiments. J. Phys. G. Nucl. Partic. 43, 045113 (2016).
[78] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., Generative adversarial networks. Commun. ACM 63, 139-144 (2020).
[79] D.P. Kingma, M. Welling, Auto-encoding variational bayes. (2013).
[80] D. Derkach, N. Kazeev, F. Ratnikov et al., Cherenkov detectors fast simulation using neural networks. Nucl. Instrum. Meth. A 952, 161804 (2020).
[81] M. Paganini, L. de Oliveira, B. Nachman, Accelerating science with generative adversarial networks: an application to 3D particle showers in multilayer calorimeters. Phys. Rev. Lett. 120, 042003 (2018).
[82] D. Salamani, S. Gadatsch, T. Golling et al., Deep generative models for fast shower simulation in ATLAS. IEEE 14th International Conference on E-Science (e-Science) 348 (2018).
[83] D. Belayneh, F. Carminati, A. Farbin et al., Calorimetry with deep learning: particle simulation and reconstruction for collider physics. Eur. Phys. J. C 80, 1-31 (2020).
[84] G.L. Ke, Z.H. Xu, J. Zhang et al., DeepGBM: a deep learning framework distilled by GBDT for online prediction tasks. Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining 384-394 (2019).
[85] S. Badirli, X.Q. Liu, Z.M. Xing et al., Gradient boosting neural networks: Grownet. (2020).
[86] W.B. He, X.G. Cao, Y.G. Ma et al., Application of EQMD model to researches of nuclear exotic structures. Nuclear Techniques 37 (2014).
[87] X.G. Cao, Y.G. Ma, Progress of theoretical and experimental studies on α cluster structures in light nuclei. Chinese. Sci. Bull. 60, 1557-1564 (2015).
[88] W.B. He, Y.G. Ma, X.G. Cao et al., Dipole oscillation modes in light α-clustering nuclei. Phys. Rev. C 94, 014301 (2016).
[89] W.B. He, Y.G. Ma, X.G. Cao et al., Giant dipole resonance as a fingerprint

of α clustering configurations in 12C and 16O. Phys. Rev. Lett. 113, 032506 (2014).

Note: Figure translations are in progress. See original paper for figures.