Transfer learning empowers material Z classification with muon tomography

Authors: Haochen Wang, Zhang, Mr. Zhao, Yu, Dr. Pei, Bao, Miss Yuxin, Zhai, Prof. Jiajia, Xu, Dr. Yu, Deng, Dr. Li, Xiao, Dr. Sa, Dr. Xueheng Zhang, Yu, Dr. Yuhong, He, Dr. Weibo, Chen, Dr. Liangwen, Zhang, Prof. Yu, Yang, Prof. Lei 杨磊, Sun, Prof. Zhiyu, Chen, Dr. Liangwen

Date: 2025-05-28T12:08:55+00:00

Abstract

Cosmic-ray muon sources exhibit distinct scattering angle distributions when interacting with materials of different atomic numbers (Z values), facilitating the identification of various Z-class materials, particularly those radioactive high-Z nuclear elements. Most of the traditional identification methods are based on complex muon event reconstruction and trajectory fitting processes. Supervised machine learning methods offer some improvement but rely heavily on prior knowledge of target materials, significantly limiting their practical applicability in detecting concealed materials. For the first time, transfer learning is introduced into the field of muon tomography in this work. We propose two lightweight neural network models for fine-tuning and adversarial transfer learning, utilizing muon tomography data of bare materials to predict the Z-class of coated materials. By employing the inverse cumulative distribution function method, more accurate scattering angle distributions could be obtained from limited data, leading to an improvement by nearly 4% in prediction accuracy compared with the traditional random sampling based training. When applied to coated materials with limited labeled or even unlabeled muon tomography data, the proposed method achieves an overall prediction accuracy exceeding 96%, with high-Z materials reaching nearly 99%. Simulation results indicate that transfer learning improves prediction accuracy by approximately 10% compared to direct prediction without transfer. This study demonstrates the effectiveness of transfer learning in overcoming the physical challenges associated with limited labeled/unlabeled data, highlights the promising potential of transfer learning in the field of muon tomography.

Full Text

Transfer Learning Empowers Material Z Classification with Muon Tomography

Haochen Wang,[1],* **Zhao Zhang,**[2,3],* **Pei Yu,**[4,2] **Yuxin Bao,**[1] **Jiajia Zhai,**[4,2] **Yu Xu,**[4,2] **Li Deng,**[4,2] **Sa Xiao,**[5] **Xueheng Zhang,**[2,6,4] **Yuhong Yu,**[2,6,4] **Weibo He,**[5],† **Liangwen Chen,**[4,2,6],‡ **Yu Zhang,**[1],§ **Lei Yang,**[2,4,6] and **Zhiyu Sun**[2,4,6]

[1]School of Physics, Hefei University of Technology, Hefei 230601, China
[2]Institute of Modern Physics, CAS, Lanzhou 730000, China
[3]Frontiers Science Center for Rare Isotopes, Lanzhou University, Lanzhou, 730000, China
[4]Advanced Energy Science and Technology Guangdong Laboratory, Huizhou 516000, China
[5]Institute of Materials, China Academy of Engineering Physics, Jiangyou 621907, China
[6]School of Nuclear Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China

*These authors contributed equally to this work.
†Corresponding author: njuyyf@163.com
‡Corresponding author: chenlw@impcas.ac.cn
§Corresponding author: dayu@hfut.edu.cn

ABSTRACT

Cosmic-ray muon sources exhibit distinct scattering angle distributions when interacting with materials of different atomic numbers (Z values), facilitating the identification of various Z-class materials, particularly radioactive high-Z nuclear elements. Most traditional identification methods rely on complex statistical iterative reconstruction or simple trajectory approximation. While supervised machine learning methods offer some improvement, they depend heavily on prior knowledge of target materials, significantly limiting their practical applicability in detecting concealed materials. For the first time, this work introduces transfer learning into the field of muon tomography. We propose two lightweight neural network models for fine-tuning and adversarial transfer learning, utilizing muon scattering data from bare materials to predict the Z-class of materials coated by typical shieldings (e.g., aluminum or polyethylene), simulating practical scenarios like cargo inspection and arms control. By introducing a novel inverse cumulative distribution-based sampling method, more accurate scattering angle distributions can be obtained from data, leading to an improvement of nearly 4% in prediction accuracy compared with traditional random sampling-based training. When applied to coated materials with limited labeled or even unlabeled muon tomography data, the proposed method

---

achieves an overall prediction accuracy exceeding 96%, with high-Z materials reaching nearly 99%. Simulation results indicate that transfer learning improves prediction accuracy by approximately 10% compared to direct prediction without transfer. This study demonstrates the effectiveness of transfer learning in overcoming physical challenges associated with limited labeled/unlabeled data and highlights the promising potential of transfer learning in muon tomography.

INTRODUCTION

Muons were discovered in cosmic rays by Carl D. Anderson and Seth H. Neddermeyer in 1937. These muons are generated when primary cosmic rays collide with atomic nuclei in the upper atmosphere, initiating nuclear-electromagnetic cascades. Most cosmic-ray muons originate from the decays of charged pions and kaons generated in these interactions. Below 1 GeV, the cosmic-ray muon energy spectrum is nearly flat, but it steepens in the energy range of 10–100 GeV, closely following the primary cosmic-ray spectrum. Above 100 GeV, the spectrum becomes even steeper because high-energy pions are more likely to interact with the atmosphere before decaying into muons [1]. At sea level, they are the most abundant charged particles, with an intensity of approximately 1 cm^{-2} min^{-1} [2, 3]. Like other charged particles, muons interact with atomic matter, leading to energy loss and multiple scattering. However, their interactions with matter are purely electroweak, resulting in significantly lower energy loss compared to most other particles, which grants them exceptional penetration capability. Consequently, over the past decades, muon tomography has demonstrated an important role in detection and imaging [4–7].

In 1970, muon transmission was first developed for discovering new chambers inside a pyramid by Alvarez et al. [8]. Since then, muon techniques have been widely applied to many fields such as nuclear safeguard [9], volcano studies [10], and underground tunneling [11]. Cheng et al. applied muon radiography to investigate the internal density distribution of the Laoheishan volcanic cone [12]. In 2003, the Los Alamos National Laboratory (LANL) first introduced the application of muon scattering tomography to security detection and material identification [13–15], underscoring its immense potential in detecting special nuclear materials such as illicit uranium concealed within cargo and containers. Leveraging the distinctive physical properties of muons, this technology has become an effective method for detecting large-scale, high-density objects.

As groupings of materials based on their atomic number, typically divided into low-Z, mid-Z, and high-Z categories, Z classification reflects both physical characteristics (e.g., scattering behavior) and practical needs in inspection and nuclear verification tasks [16]. Z-class identification of materials based on muon scattering is a crucial task for security screening and industrial applications.

Xiao et al. presented a modified multi-group model to improve the image resolution of high-Z materials [17]. Ji et al. proposed a method for imaging materials using the ratio of secondary particles produced by muons [18]. A novel imaging reconstruction method with big voxel and angle capping was proposed to reduce time and storage consumption at the Institute of Modern Physics, CAS [19]. While most traditional identification methods rely on complicated reconstruction of the muon event and track fitting process, which significantly increases the design and calculation cost of the algorithm.

Compared to traditional physics-based reconstruction methods, deep learning models can automatically extract and learn complex features from data. This end-to-end learning paradigm significantly improves computational accuracy and efficiency [20, 21]. Deep learning techniques have demonstrated outstanding performance in various fields, including image reconstruction [22–24], nuclear physics, and particle physics [25–27]. Common deep learning methods rely on supervised learning, which is based on labeled samples for training. Gao et al. proposed a convolutional neural network model for feature extraction to realize the classification and recognition of materials based on muon scattering [28]. Our previous work also explored a feasible solution for muon-based material identification using supervised deep learning [29, 30]. However, in practical identification scenarios, obtaining labeled scattering angle data for coated materials is the major challenge. The scarcity of labeled data limits the practical application of supervised learning in coated material prediction.

As a critical deep learning strategy, transfer learning enables knowledge learned from one domain (the source domain) to be transferred to another related domain (the target domain), effectively mitigating the problem of limited labeled data in the target domain [31–33]. This inspires a proposal of a lightweight neural network model based on transfer learning for Z-class identification of coated materials using muon scattering data. In this study, we define the Z-class identification of bare materials as the source domain task and that of coated materials as the target domain task. This formulation is inspired by realistic application scenarios—such as cargo inspection [34], nuclear safeguard verification, and arms control [35, 36]—where the internal composition is unknown or only partially known. By introducing a novel data preprocessing and sampling method, the model transfers the feature-label mapping learned from bare material scattering angles to the Z-class prediction of coated materials. We designed two coated material scenarios, where Al and PE served as the coating materials (Fig. 1 [Figure 1: see original paper]). Utilizing two transfer learning paradigms, fine-tuning [37] and Domain-Adversarial Neural Network (DANN) [38], we achieve Z category classification for nine coated materials, which include three materials from each of the high, mid, and low-Z categories.

To evaluate the effectiveness of our proposed approach, we conducted a series of simulations on the Z-class classification task with muon scattering angle data using Geant4 Monte Carlo simulation [39]. The results demonstrate that this approach effectively transfers scattering angle features from bare materials, en-

abling accurate classification of coated ones. Our method not only reduces reliance on large-scale labeled data for coated materials but also maintains excellent Z-class classification accuracy in a few minutes, particularly achieving superior accuracy in the more application-critical high-Z class identification. The contributions of our work lie in the following three aspects:

1. Proposing a novel sampling method combining inverse cumulative distribution function (CDF), integration, and interpolation that improves the feature expression ability of muon scattering angle samples, optimizing the training effect of neural networks.

2. Development of a fine-tune based transfer learning model for fast Z-class prediction in cases of scarce coated material labels, and a DANN transfer learning model based on adversarial ideas for stable Z-class prediction when coated material labels are completely unknown.

3. Comprehensive result analysis and physical correlation interpretation, demonstrating the potential application scenarios of each method, which provides a diversified scheme for the application and expansion of transfer learning in muon techniques such as cargo inspection and nuclear safeguard.

To the best of our knowledge, this is the first study to apply transfer learning strategies to the field of muon tomography. This research provides an alternative machine learning scheme for traditional identification methods. It demonstrates the great potential of transfer learning in mitigating the high-cost reconstruction and data scarcity challenges in muon-based applications with similar scenarios.

Fig. 1: Flow diagram of material Z classification with transfer learning. The bare material is defined as the source domain, while the coated material is defined as the target domain.

II. MUON SCATTERING SIMULATION AND SAMPLING METHOD

A. Simulation Setups

The scattering angle data are provided by Geant4 simulation incorporating the CRY (Cosmic-Ray Shower Library) software [40], which generates muons with the energy and angular distribution of cosmic muons at sea level. The dimension of the object is 10 cm × 10 cm × 10 cm; it may be coated on all six sides with a 1 cm thick layer, resulting in an overall size of 12 cm × 12 cm × 12 cm for the coated object. Four position-sensitive detectors of size 30 cm × 30 cm are placed above and below the object to measure the trajectories of the incident and emergent muons for subsequent angle calculations. A spacing of 20 cm is placed between the second and third detector for positioning the tested materials at the center of this gap. The distance between detector 1, 2 and detector 3, 4 is

the same 35 mm. It is worth noting that the detector used in the simulation is an ideal detector, and detector characteristics like detector noise are not taken into account. This setup is illustrated in Fig. 2 [Figure 2: see original paper].

Fig. 2: Schematic of the Geant4 simulation setup.

First, we set up a muon source with an energy of 1 GeV and conducted scattering angle simulations for nine bare materials (Mg, Al, Ti, Fe, Cu, Zn, W, Pb, U) to verify the reliability of the simulated data. Fig. 3 [Figure 3: see original paper] presents the probability distribution statistics of the simulated scattering angle data using kernel density estimation (KDE), categorized by both Z classes and different materials. The results indicate significant differences in the scattering angle distributions among different Z-class materials, as well as among materials within the same Z class. In the formal experiment, we performed simulations based on the energy and angular distributions of cosmic-ray muons at sea level. For each material in different scenarios (bare, Al coated, or PE coated), 500,000 muon scattering angle data were simulated. The scattering angle in our simulation is the included angle between incoming and outgoing muon tracks in 3D space using the formula:

$$\theta = \arccos \left(\frac{\vec{\mu}_{in} \cdot \vec{\mu}_{out}}{|\vec{\mu}_{in}||\vec{\mu}_{out}|} \right)$$

B. Inverse CDF Sampling

As shown in Fig. 3, the primary distinguishing feature of muon scattering through different materials lies in the variations of their scattering angle distributions. Additionally, high-quality training samples contribute to the training of more effective neural networks [41, 42]. When employing a neural network as a mapping function between scattering angle data and material properties, it is desirable for the model to learn features that facilitate differentiation between various materials based on their respective scattering characteristics. To effectively capture the distribution characteristics of the simulation data, we introduce a sampling method called the inverse cumulative distribution function. First, we uniformly select a set of quantile points within the range of the overall scattering angle distribution for a given material and use non-parametric methods to estimate the probability density function (PDF) of the overall data at these quantile points. The corresponding CDF values are then computed based on these PDF values. Finally, we obtain training samples through inverse CDF sampling. The complete sampling process is shown in Fig. 4 [Figure 4: see original paper], and the process pseudocode is referred to in Appendix 1. Compared with the traditional random sampling (RS) method, the samples generated using this method share the same probability distribution as the overall simulation data.

We employed two non-parametric estimation methods: KDE [43] and histogram estimation (HE), to compute the probability density values for selected quantile

points within the overall data. The fundamental idea of KDE is to use a smooth kernel function to perform weighted averaging of total data points, thereby obtaining an estimate of the probability density. In other words, we calculate the weighted contribution of data points at a given quantile point. Selecting specific quantile points rather than drawing the entire data ensures computational efficiency and accurate distribution estimation, mitigating the impact of fine-grained details that could disrupt the smoothness of the overall probability distribution. For a selected quantile point X_q, its kernel density estimation of PDF is given by:

$$f_k(X_q) = \sum K\left(\frac{X_q - X_i}{h}\right)$$

where K is the kernel function, h is the bandwidth parameter controlling the smoothness of the estimation, and X_i are the true data values in the overall data. We employ a Gaussian function as the kernel function, along with the Silverman method for automatic bandwidth adjustment.

In contrast, the HE method is more straightforward and intuitive. It uniformly divides the overall data range into b bins and counts the number of data points falling within each bin:

$$f_h(x) = \sum \frac{1(X_i \in B_j)}{nh}$$

where h is the bin width, n is the total number of samples. The expression $1(X_i \in B_j)$ represents an indicator function, which indicates whether the data point X_i falls within the bin B_j. If so, the function takes the value of 1, otherwise it takes the value of 0. The KDE method provides a smoother estimation of the PDF values, enabling the subsequent acquisition of higher-quality samples. However, for sparse data or the presence of extreme values, KDE may lose some accuracy. On the other hand, the HE method is more stable, but improper bin selection can lead to overfitting or underfitting of the data distribution. Given the peak and long-tailed distribution characteristics of the simulation data, we incorporate both methods into the sampling process.

To convert discrete PDF values into continuous CDF values, the compound trapezoidal rule (Eq. 4) and cubic spline interpolation (Eq. 5) are employed during the converting process. Dividing the integral interval into multiple subintervals and applying the trapezoidal rule within each one improves integration accuracy:

$$F(X_q) \approx \frac{X_{q+1} - X_q}{2}\left[f(X_1) + 2\sum_{p=2}^{q-1} f(X_p) + f(X_q)\right]$$

where $f(X_q)$ is obtained from Eq. 2 or Eq. 3, depending on which method of calculating PDF is taken. A smooth and continuous CDF is obtained by using cubic spline interpolation between discrete CDF values, and the parameters a_j, b_j, and c_j are determined by zero, one, and two-order boundary conditions for different pairs of CDF values:

$$F_j(x) = a_j(x - X_i)^3 + b_j(x - X_i)^2 + c_j(x - X_i) + d_i$$

Based on the CDF values obtained from the above calculations and converting, N_s rounds of sampling are conducted for various materials' simulation data under different experimental scenarios. In each iteration, n random numbers u_n that satisfy uniform distribution $U \sim (0,1)$ are generated, and the inverse CDF method is utilized to obtain n scattering angle data points $\{x_1, x_2, ..., x_n\}$:

$$x_i = F^{-1}(u_i)$$

To prevent overfitting and a loss of robustness due to high similarity between samples, a similarity check mechanism is incorporated. By calculating the Euclidean distance between samples generated in each iteration, those with excessive similarity are filtered out, ultimately resulting in N_s diverse samples that align with the overall simulation data distribution. The parameters related to the sampling process are set in Table 1 .

TABLE 1 : Parameter setting of sampling process.

Parameter	Value	Description
tot mat sim data	500,000	Total number of materials
smp num (Ns)	1,000	Simulated muon scattering data for each material
scat num (n)	1,000	The number of samples generated for a material
qtl point	1,000	The number of muon scattering data contained in each sample
bin num	1,000	The number of quantile points where KDE calculating PDF values
Silverman	-	Bandwidth adjustment mode in
sim thd	5×10^{-3}	The threshold of similarity discrimination between samples

In different scenarios (bare, Al coated, or PE coated), 1,000 samples were acquired for each material. While in specific training and prediction phases, a

consistent train-test split will be applied across different materials. The specific data partitioning ratios for various phases will be detailed in the corresponding parameter tables. In addition, the traditional random sampling method, which directly samples from the raw simulated scattering angle data and is denoted as RS, is also applied to the study. The sample size generated by the RS method is the same as that of inverse CDF sampling. According to the comparison of training accuracy, the superiority of the inverse CDF sampling method can be demonstrated.

III. TRANSFER LEARNING-BASED Z-CLASS IDENTI-FICATION

There are some implicit correlation features between source domain tasks and target domain tasks, which constitutes the practical feasibility basis of transfer learning [31–33]. In this study, we adopt two transfer learning paradigms: fine-tuning learning and adversarial transfer learning with DANN. Based on a pretrained model trained in the source domain, fine-tuning performs limited parameter adaptation in the target domain and enables efficient learning of feature-label relationships even when target domain data is scarce. Adversarial transfer learning, on the other hand, is applicable when target domain labels are completely unknown. By extracting shared discriminative features, it aligns the feature distributions between the source and target domains, enabling classification in an unsupervised manner.

A. Pre-training and Fine-tuned Transfer

As an essential technique for transferring neural network tasks, fine-tuning is widely applied in transfer learning due to its low computational cost and high training efficiency under limited target domain data conditions. In this study, we constructed a unified lightweight neural network with two hidden layers for both pre-training and fine-tuning processes (P&F model). The scattering angle sample data is first received by the input layer, then processed through two hidden layers for feature extraction, and finally classified by the output layer. The detailed network structure is shown in Fig. 5 [Figure 5: see original paper].

TABLE 2 : Parameter setting in pre-training & fine-tuning task. Unless specified with asterisks, certain parameters are shared between the pre-training and fine-tuning stages.

Parameter	Value	Description
input dim	1000	The number of nodes in the P&F model input layer
hidden dim1	256	The number of nodes in the P&F model first hidden layer

Parameter	Value	Description
hidden dim2	256	The number of nodes in the P&F model second hidden layer
output dim	3	The number of nodes in the P&F model output layer
tt ratiop	0.3	Ratio of source samples in training set to test set during pre-training
tt ratiof	0.3	Ratio of target samples in training set to test set during fine-tuning
batch size	64	The number of samples trained for each material in one iteration
epochp	50	The number of epochs in the pre-training process
epochf	50	The number of epochs in the fine-tuning process
lr	0.001	The learning rate of the P&F model in the training process

Since the P&F model is designed for multi-classification tasks, we use cross-entropy as the loss function [44]:

$$L_{P\&F} = -\sum_{i=1}^{N} \sum_{j=1}^{C} y_{i,j} \log(\text{Softmax}(z_{i,j}))$$

where N is the batch-size and C is the number of classes (low, mid, and high). Given a sample i, the ground-truth class label $y_{i,j}$ is encoded in a One-Hot format and automatically converted by PyTorch. The corresponding logits $z_{i,j}$ represent the raw predictions of the neural network, which will be normalized by the Softmax activation function. The model is trained on the training set, and its prediction accuracy on the test set is recorded to evaluate its generalization performance and robustness. The specific parameter settings are detailed in Table 2.

TABLE 3 : Classification accuracies of pre-trained model on source dataset. The "Total" column represents overall prediction accuracies across three categories.

Sampling Method	Z categories	Low-Z	Mid-Z	High-Z	Total
RS		0.923	0.947	0.967	0.946
KDE		0.967	0.987	0.993	0.983

Sampling Method	Z categories	Low-Z	Mid-Z	High-Z	Total
HE		0.960	0.983	0.990	0.980

For the pre-training process, we trained the P&F model with feature-label pair samples (x_s, y_s) obtained from the source domain through three different sampling methods. The detailed training process is shown in Fig. 6 [Figure 6: see original paper]. The goal is to learn the mapping between the scattering angle data of bare materials and their Z categories. After pre-training, the hidden layers, which serve as the key structures for feature extraction, have their parameters optimized to effectively extract high-order features from the original scattering angle data. In the subsequent fine-tuning process, the parameters of the input and hidden layers will be frozen, while training only the parameters between the last hidden layer and the output layer using a smaller number of target domain training samples.

In the pre-training stage, we aimed at nine different materials from the high, mid, and low-Z categories and conducted supervised training on the training dataset. The training results in Table 3 indicate that the pre-trained model achieves high prediction accuracy on the test dataset, confirming the effectiveness of neural networks in learning the mapping between scattering angle data and Z categories. The accuracy metric is the ratio of correct predictions to total sample number in their respective scenarios.

Before applying the fine-tuned transfer learning method, we first directly evaluated the pre-trained model on the two target domains where Al and PE serve as coating materials. The test results are shown in Table 4 (Pre-train). Under training with samples obtained using two inverse CDF sampling methods (KDE and HE), the classification accuracy in the Al-coated target domain decreased by approximately 12%, with the most significant drop occurring in the prediction accuracy of low-Z materials (about 30%). This phenomenon can be attributed to the fact that, as a low-Z material, the relative influence of Al coating on the scattering angle distribution of the coated material is inversely proportional to the intrinsic Z value of the coated material. Meanwhile, in the PE-coated target domain, the pre-trained model's prediction accuracy remained almost unchanged (a decrease of only about 1%). Since PE, as a hydrocarbon compound, can be considered an ultra-low-Z material, its coating has minimal impact on the scattering angle distribution of metallic materials. In contrast, when the model was trained with RS data, the randomness of the sampling process hindered the prediction results from following the physically consistent patterns observed in inverse CDF-based training.

In our P&F model architecture, the first hidden layer is designed to extract fundamental features reflecting low-level scattering physics. These features are largely invariant across domains, given the shared nature of the input data structure between source and target domains. Therefore, freezing the first hidden

layer is intended to preserve these transferable, physically meaningful representations. Regarding the second hidden layer, although it captures more domain-specific cues, our empirical studies showed that fine-tuning this layer—while the first remains frozen—often leads to overfitting due to the limited size of the target dataset. On the other hand, freezing both hidden layers stabilizes performance by retaining shared representations and reducing the influence of domain-specific noise. Additionally, freezing both layers reduces the number of trainable parameters, which is especially beneficial in low-cost scenarios.

For the fine-tuning process, we freeze the corresponding parameters and fine-tune the P&F network with a smaller number of target domain training samples (x_t, y_t). This enables the pre-trained model to be adapted to target domain tasks efficiently. For the two different target domains of Al-coated and PE-coated materials, the classification accuracy of the fine-tuned model on the target test dataset is presented in Table 4 (Fine-tune). Benefiting from the well-optimized parameters obtained during pre-training, the fine-tuned P&F model demonstrates excellent predictive performance across both tasks. The detailed training process is shown in Fig. 7 [Figure 7: see original paper].

TABLE 4: Prediction accuracies on target dataset before/after fine-tune transfer.

Training Stage	Dataset	Sampling Method	Z categories	Low-Z	Mid-Z	High-Z	Total
Pre-train	Al-coated	RS		0.617	0.853	0.967	0.813
		KDE		0.700	0.880	0.987	0.857
		HE		0.693	0.877	0.987	0.853
Fine-tune	Al-coated	RS		0.843	0.957	0.993	0.933
		KDE		0.933	0.987	1.000	0.973
		HE		0.930	0.987	1.000	0.973
Pre-train	PE-coated	RS		0.930	0.967	0.987	0.960
		KDE		0.967	0.987	0.997	0.983
		HE		0.967	0.987	0.997	0.983
Fine-tune	PE-coated	RS		0.967	0.987	0.997	0.983
		KDE		0.987	0.997	1.000	0.993
		HE		0.987	0.997	1.000	0.993

It should be noted that, due to the globally shared parameters of the neural network, the fine-tuning process aims to enhance overall classification accuracy rather than optimizing each class individually. As the model improves its discriminative ability for certain categories, performance on others may deteriorate slightly, resulting in parameter competition and trade-offs. This phenomenon may cause a minor decrease in prediction accuracy for specific classes. However, the overall classification performance on the target task still improves. Moreover, the observation that the training loss exceeds the testing loss, with the training curve showing greater fluctuation while the test curve remains smooth,

can be primarily attributed to the limited size of the training set. Since only 30% of the total data is used for training, the model becomes more sensitive to outliers and difficult samples, leading to higher average loss and increased variance during optimization. In contrast, the larger testing set (70%) provides a more reliable and representative performance estimate, effectively averaging out anomalies and resulting in a smoother and lower loss curve. Furthermore, as parameter updates occur solely on the training set, its loss is expected to fluctuate across batch iterations, whereas the testing loss remains relatively stable.

B. Transfer Learning with DANN Model

The identification of the Z-class of an unknown coated material, as an unlabeled target domain problem, presents a more significant challenge. However, by identifying common scattering angle features between the coated material and its bare counterpart, we can train a neural network to achieve superior classification performance in the unknown domain. Traditional domain alignment methods typically compute specific mathematical relationships between the source and target domains and incorporate them into the training process as part of the loss function [45, 46]. However, given that different transfer learning tasks exhibit distinct data characteristics, determining the optimal mathematical relationship as a training objective remains a highly challenging task.

The concept of adversary in neural networks was first introduced in [47], where adversarial models generate adversarial samples to enhance model robustness. The DANN extends adversarial training to transfer learning by incorporating a domain discriminator that enforces feature distribution alignment between the source and target domains through adversarial training, thereby facilitating unsupervised learning in the target domain. This approach fully exploits the fitting capabilities of neural networks and enables effective feature alignment without requiring an explicit definition of the feature relationships between a specific source and target domain.

Our DANN model, as illustrated in Fig. 8 [Figure 8: see original paper], consists of three main components: a feature extractor, a classifier, and a domain discriminator. During training, the labeled scattering angle data of bare materials and the unlabeled scattering angle data of coated materials are both fed into the feature extractor. The total extracted features are then passed into the domain discriminator, while the extracted source domain features and corresponding labels serve as input to the classifier.

First, the domain discriminator determines whether the input features originate from the source or target domain and simultaneously computes the loss function L_{disc} to optimize itself, aiming to improve domain feature discrimination. Since the domain discriminator only classifies whether the received features belong to the source domain or the target domain, we use binary cross-entropy as the loss function:

$$L_{disc} = -\sum_{i=1}^{N} \left[y_i \log(\text{Sigmoid}(-z_i)) + (1 - y_i) \log(1 - \text{Sigmoid}(-z_i)) \right]$$

where the ground-truth label $y_i \in \{0, 1\}$ indicates the source domain class or the target domain class. The model outputs logits z_i representing the raw predictions before applying the Sigmoid activation function. Unlike the default definition for Softmax in cross-entropy, the binary cross-entropy loss requires explicitly defining Sigmoid as the activation function in the output layer.

Second, when receiving a sample $(F(x_s), y_s)$ from the source domain, the classifier computes the loss function L_{cls} for the conventional multi-classification task, similar to the pre-training and fine-tuning approaches. Consequently, the loss function L_{cls} is also defined as cross-entropy, with the same expression as Eq. 7. Finally, the overall loss function:

$$L_{total} = L_{cls} - \lambda L_{disc}$$

is used to train the extractor, where L_{cls} is responsible for improving the prediction accuracy of the extractor and classifier, while the reversal discrimination loss $-\lambda L_{disc}$ is used to train the extractor in a way that gradually extracts shared features between the source and target domains, thereby achieving domain alignment. Gradient reversal ensures that the extractor and discriminator form an adversarial relationship during training, eventually reaching equilibrium after iterative optimization. The gradient reversal parameter λ is employed to balance the weight distribution between feature alignment and classification accuracy improvement in the network model. In this study, feature alignment is more challenging than classification. Moreover, due to the minimal feature distribution discrepancy between the source and target domains, L_{disc} remains relatively low. Based on this analysis, we assign a higher weight to λ in our training process to enhance training performance. After this training stage, the extractor and classifier possess the ability to extract common features from both the source and target domains and perform effective classification. Detailed neural network hyperparameter settings are shown in Table 5.

TABLE 5 : Parameter setting of DANN transfer. Considering DANN as an integrated model, the input layers of both the classifier and the discriminator correspond to the hidden layers of the entire network.

Parameter	Value	Description
input dimf	1000	The number of nodes in the feature extractor input layer
hidden dimf1	256	The number of nodes in the feature extractor first hidden layer

Parameter	Value	Description
hidden dimf2	256	The number of nodes in the feature extractor second hidden layer
hidden dimc	256	The number of nodes in the classifier input layer
output dimc	3	The number of nodes in the classifier output layer
hidden dimd1	256	The number of nodes in the discriminator input layer
hidden dimd2	256	The number of nodes in the discriminator hidden layer
output dimd	1	The number of nodes in the discriminator output layer
tt ratio	0.3	Ratio of samples in training set to test set
batch size	64	The number of samples trained simultaneously in one iteration
epoch	50	The number of epochs in the training process
lrf	0.001	The learning rate of the feature extractor in the training process
lrc	0.001	The learning rate of the classifier in the training process
lrd	0.001	The learning rate of the discriminator in the training process
grad rev (λ)	0.5	Parameters for gradient reversal

The training process and final results of DANN are presented in Fig. 9 [Figure 9: see original paper] and Table 6 . Training results indicate that, when trained with inverse CDF sampled data, DANN exhibits slightly lower prediction accuracy for low-Z materials compared to mid and high-Z materials. This phenomenon is consistent with the results obtained when the pre-trained model performs inference directly in the target domain, which can be attributed to the significant change in the scattering angle distribution of low-Z materials after being coated in Al, making feature alignment more challenging than for mid-Z and high-Z materials. Meanwhile, the introduction of a large gradient reversal coefficient $-\lambda$ causes the overall training loss to become negative, as it includes the adversarial loss from the domain discriminator. However, this

does not affect the evaluation on the test set, where the domain discriminator and gradient reversal are not involved. Therefore, the test loss, calculated solely based on the standard cross-entropy loss, remains positive. This behavior is expected and does not impact the assessment of model performance on the target classification task.

TABLE 6: Prediction accuracies on target dataset after DANN transfer.

Dataset	Sampling Method	Z categories	Low-Z	Mid-Z	High-Z	Total
Al-coated	RS		0.800	0.940	0.993	0.913
	KDE		0.900	0.980	1.000	0.967
	HE		0.897	0.980	1.000	0.967
PE-coated	RS		0.967	0.987	0.997	0.983
	KDE		0.987	0.997	1.000	0.993
	HE		0.987	0.997	1.000	0.993

IV. RESULTS AND DISCUSSION

The overall prediction accuracy of the pre-trained model, the fine-tuned model, and the DANN model in the target domain is summarized in Table 7 . The training results indicate that the introduction of the inverse CDF sampling method effectively improves sample quality, thereby enhancing prediction accuracy. Due to the randomness of the RS method and the inherent black-box nature of neural networks, with RS methods, it is difficult to provide a clear physical explanation for the variation of the training results in the source and target domain. However, since the inverse CDF sampling method effectively captures the scattering angle data distribution, it offers stronger interpretability to the result. Additionally, as stated in [48], a total of 1,400 scattering instances is sufficient to construct a statistically reliable muon scattering angle probability distribution. Although we compute the CDF at selected data points, our global interpolation acts on a total of 500,000 instances, which is far beyond this threshold. Therefore, in practical applications, as long as sample diversity is maintained, the total amount of training data required can be further reduced.

For the two target domain tasks we considered, since PE coating has a minimal impact on the muon scattering angle distribution, the prediction accuracy of the model in the PE task only slightly improved (close to 100%) after transfer learning. However, in the Al task, transfer learning improves prediction accuracy by approximately 10%. When comparing the two transfer learning methods, fine-tuning and DANN, the training results show that the fine-tuned model achieves slightly higher prediction accuracy than the DANN model. As seen in the previously summarized results in Section III, fine-tuning benefits from the supervised learning process, leading to more balanced prediction accuracy across different Z-class materials. In contrast, although DANN improves the prediction

accuracy of low-Z materials by more than 20% compared to the pre-trained model without transfer learning, its accuracy remains slightly lower than that of fine-tuning due to the unsupervised training in the target domain. However, this also highlights one of DANN's key advantages: it does not require any labels from the target domain, yet it achieves prediction accuracy comparable to fine-tuning. Moreover, since the DANN-based transfer learning approach focuses on extracting shared features between the source and target domains, whereas fine-tuning involves learning the entire feature space of the source domain before transferring to the target domain, this may capture irrelevant features that do not contribute to the target domain task. As a result, the training process indicates that DANN achieves greater stability and robustness compared to fine-tuning. This makes DANN particularly valuable for real-world applications where fully unsupervised transfer learning is often required.

Parameters of the two models are in the $10^5 \sim 10^6$ magnitude (P&F model contains 72,579 parameters and DANN model contains 323,332 parameters). In our implementation, the two transfer learning methods require running memory on the order of $10^2 \sim 10^3$ MB and run in a few minutes on a local Intel Core i9-14900K (24-core, 32-thread, up to 6.0 GHz) computer. The model architectures, training process, and data loading were implemented using PyTorch libraries. The above analysis can fully prove that our model and algorithm have relatively low deployment and training costs.

TABLE 7: Prediction accuracies on total target dataset (all Z-categories) with different training methods. Pre-training is a direct prediction without transfer.

Training Method	Dataset	Sampling Method	Total
Pre-train	Al-coated	RS	0.813
		KDE	0.857
		HE	0.853
Fine-tune	Al-coated	RS	0.933
		KDE	0.973
		HE	0.973
DANN	Al-coated	RS	0.913
		KDE	0.967
		HE	0.967
Pre-train	PE-coated	RS	0.960
		KDE	0.983
		HE	0.983
Fine-tune	PE-coated	RS	0.983
		KDE	0.993
		HE	0.993
DANN	PE-coated	RS	0.983
		KDE	0.993
		HE	0.993

V. CONCLUSION

We developed a novel transfer learning method for identifying material Z-class using muon scattering angle data, which serves as an alternative to traditional identification methods based on complex physical model reconstruction. First, Monte Carlo simulations were conducted using Geant4 to obtain scattering angle data for specified materials in both bare and coated states. A series of fitting techniques, including computation and interpolation, were employed to derive the probability distribution of material scattering angles. Based on this distribution, we generated a sampled dataset that better conforms to the overall distribution, resulting in an approximately 4% improvement in prediction accuracy in the target domain and enhancing the physical interpretability of the training process.

Meanwhile, in real-world application scenarios, the correspondence between scattering angle data and the coated material is often unknown. To address this challenge, we introduced two novel lightweight neural networks trained using transfer learning. By employing either fine-tuned supervised learning or adversarial unsupervised learning on the coated material, these models transfer source-domain knowledge learned from bare material data to the target domain of coated materials. In the PE target domain task, where the scattering angle distribution remains largely unchanged before and after coating, the prediction accuracy reaches 99%. In contrast, for the more challenging Al-coated task, the prediction accuracy improves by approximately 10% compared to the pre-transfer learning model. The results demonstrate that our method achieves high prediction accuracy even when the mapping between coated material and scattering data is scarce or completely unknown.

We analyzed the results under different tasks and scenarios, verifying that Z-class identification based on machine learning aligns with the physical principles of muon interactions, validating the feasibility of Z-class prediction for coated materials via transfer learning. Furthermore, this study reveals that features learned from data through machine learning exhibit transferability, rather than merely relying on the repeated application of domain-specific expertise across different scenarios. This suggests that machine learning methods based on transfer learning can serve as a cost-effective training approach for conducting physics research tasks in similar situations.

In future research, incorporating additional physical variables such as muon momentum beyond the scattering angle into the training data is expected to further improve the accuracy and robustness of Z classification. Real-world detection challenges include detector noise and electronic fluctuations that bias scattering angle measurements, resolution limitations that reduce the model's ability to distinguish materials with similar scattering properties, and system inefficiencies—such as dead time and low detection efficiency—that degrade

overall data quality. We may introduce some noise-aware training to boost the model's robustness and generalization in real environments. Furthermore, by introducing more statistics and enhancing the generalization ability of the model, transfer learning methods are expected to be extended to Z-value identification tasks in more complex scenarios.

ACKNOWLEDGMENTS

APPENDIX: ALGORITHMIC PSEUDO-CODE

Algorithm 1: Data Conversion and Sampling Process

Input: total material M, Z-class dictionary of materials Z, total simulation data .root

Output: training dataset D_{train} ($M \times N_s \times n \times r_1$), test dataset D_{test} ($M \times N_s \times n \times r_2$)

Parameters: number of simulation data for each material N^*; number of samples N_s, number of scattering angle n in each sample; similarity threshold k, number of mean quantile points in total data q; ratio of training dataset to test dataset $r_1 : r_2$.

```
// data format conversion
// time complexity: O(N*) space complexity: O(N*)
1. for material m in M do
2.   read .root data for material m
3.   if .root data is not None then
4.     remove NaN values and trim to same size
5.     (features, label) $\leftarrow$ (processed .root, Z-class of the material)
6.     total dataset table D*_M $\leftarrow$ (features, label) with .csv format
7.   end
8. return D*

// sampling process
// time complexity: O(m $\times$ (N* + N$^{2}$ $\times$ n)) space complexity: O(m $\times$ N
9. for data of material m D*_m in D* do
10.  X $\leftarrow$ features in D*_m
11.  Y $\leftarrow$ label in D*_m
12.  X_q $\leftarrow$ selected q point in X
```

```
13.  f(X_q) $\leftarrow$ PDF values calculated on X_q
14.  CDF with Composite Trapezoidal Rule: F(X) $\leftarrow$ (X_{q+1} - X_q)/2 $\times$ [f(X_
15.  cubic spline interpolation
16.  for I $\leftarrow$ 1 to N_s do
17.    build sample list D_m $\leftarrow$   for material m
18.    u_n $\leftarrow$ n numbers satisfy U   (0, 1)
19.    x_i $\leftarrow$ F^{-1}(u_i), where i = 1, 2, ..., n
20.    x_I $\leftarrow$ {x_1, ..., x_n}
21.    if D_m is not empty and similarity between (x_I, y) and samples in D_m > k then
22.      I $\leftarrow$ I - 1
23.    add (x_1, y) to D_m
24.  return D_m = {(x_1, y), ..., (x_{N_s}, y)}
25. end
26. integrate all material sample lists into dataset D_M
27. divide the training dataset and test dataset: D_{train} $\leftarrow$ D_M $\times$ r_1, D
28. return D_{train} and D_{test}
```

Algorithm 2: Pre-training and Fine-tune Transfer Learning

Input: source training dataset D^S_{train}, target training dataset D^T_{train}, source test dataset D^S_{test}, target test dataset D^T_{test}

Output: pre-trained model $\Theta_P(x, \theta_p)$, fine-tuned model $\Theta_F(x, \theta_f)$

Parameters: batch-size N, number of epoch T, learning rate α, optimizer Adam, neural network model $\Theta(x, \theta)$

```
// Pre-training process
1. for t $\leftarrow$ 1 to T do
2.   // Iterate over dataset D_{train}^S with batch-size N
3.   for D_{train} in D_{train}^S do
4.     X_i $\leftarrow$ labels in D_{train}
5.     Y_i $\leftarrow$ features in D_{train}
6.     Ŷ_i $\leftarrow$ θ(X_i, )
7.     loss function CrossEntropy: L = -1/N  _{j=1}^3 Y_{i,j} log Ŷ_{i,j}, where j is Z cate
8.     loss.backward()
9.      _p $\leftarrow$ update model weights   with Adam in $\alpha$
10.    acc_{train} $\leftarrow$ cumulate accuracies with Y_i and Ŷ_i
11.  end
12.  for D_{test}^N in D_{test}^S do
13.    X_i* $\leftarrow$ labels in D_{test}
14.    Y_i* $\leftarrow$ features in D_{test}
15.    Ŷ_i* $\leftarrow$ θ(X_i*, )
16.    acc_{test} $\leftarrow$ cumulate accuracies with Y_i* and Ŷ_i*
17.    (acc_1, acc_2, acc_3) $\leftarrow$ cumulate accuracies of (Low-Z, Mid-Z, High-Z) with
18.  end
19.  store acc_{train}, acc_{test}, acc_1, acc_2, acc_3
20. return pre-trained model θ_P(x,  _p)
```

```
// Fine-tuning process
21. load model θ_P(x, _p)
22. θ_F(x, ) $\leftarrow$ update only the last fully connected layer of θ_P(x, _p)
23. train model θ_F(x, ) with D_{train}^T and test with D_{test}^T (same training process a
24. store acc_{train}, acc_{test}, acc_1, acc_2, acc_3
25. return fine-tuned model θ_F(x, _f)
```

Algorithm 3: Training Process of DANN

Input: source training dataset D_{train}^S, target training dataset D_{train}^T, target test dataset D_{test}^T

Output: trained feature extractor $F(x, f)$, classifier $C(x, c)$, and domain discriminator $D(x, d)$

Parameters: batch-size N, number of epochs T, learning rate α_1, α_2, inversion parameter λ; feature extractor F, classifier C, domain discriminator D

```
// Training process of DANN
1. for t $\leftarrow$ 1 to T do
2.    for D_{train}^N from D_{train}^S and D_{test}^N from D_{train}^T do
3.       (X_S, Y_S), (X_T,  ) $\leftarrow$ D_{train}
4.       F_S $\leftarrow$ F(X_S), F_T $\leftarrow$ F(X_T)
5.       Ỹ_S $\leftarrow$ C(F_S)
6.       cls_{loss} $\leftarrow$ CrossEntropy(Ỹ_S, Y_S)
7.
8.       // Update domain discriminator
9.       domain_{labels} $\leftarrow$ [1 for F_S, 0 for F_T]
10.      domain_{pred} $\leftarrow$ D([F_S, F_T])
11.      disc_{loss} $\leftarrow$ BinaryCrossEntropy(domain_{pred}, domain_{labels})
12.      disc_{loss}.backward()
13.      d* $\leftarrow$ update D(x, d) weights d with Adam in $\alpha$_2
14.
15.      // Update feature extractor and classifier
16.      total_{loss} $\leftarrow$ cls_{loss} - $\lambda$ $\times$ disc_{loss}
17.      total_{loss}.backward()
18.      f*, c* $\leftarrow$ update F(x, f) and C(x, c) weights f and c with Adam in $\alpha$_
19.      acc_{train} $\leftarrow$ cumulate accuracies with Ỹ_S and Y_S
20.   end
21.
22.   // Evaluation on target domain
23.   set every model to evaluation mode (disable gradient updating)
24.   for D_{test}^N* from D_{test}^T do
25.      (X, Y) $\leftarrow$ D_{test}
26.      F_T $\leftarrow$ F(X)
27.      Ỹ $\leftarrow$ C(F_T)
28.      acc_{train} $\leftarrow$ cumulate accuracies with Ỹ and Y
29.      compute accuracies for each category (Low-Z, Mid-Z, High-Z)
30.   end
```

```
31.  store acc_{train}, acc_{test}, accuracy for each category (low-Z, mid-Z, high-Z)
32.  return trained model F(x, f*), C(x, c*) and D(x, d*)
```

REFERENCES

[1] P. Shukla, S. Sankrith, *Energy and angular distributions of atmospheric muons at the Earth* (2016). arXiv:1606.06907

[2] S.H. Neddermeyer, C.D. Anderson, *Cosmic-ray particles intermediate mass.* Phys. Rev. 54, 88-89 (1938). doi:10.1103/PhysRev.54.88.2

[3] M. Tanabashi, P.D. Grp, K. Hagiwara et al., *Review of particle physics: particle data group.* Phys. Rev. D. 98, 030001 (2018). doi:10.1103/PhysRevD.98.030001

[4] N. Su, Y.Y. Liu, L. Wang et al., *A Comparison of Muon Flux Models at Sea Level for Muon Imaging and Low Background Experiments.* Front. Energy. Res. 9, 750159 (2021). doi:10.3389/fenrg.2021.750159

[5] H. Miyadera, C.L. Morris, *Muon scattering tomography: review.* Appl. Opt. 61, C154-C161 (2022). doi:10.1364/AO.445806

[6] A. Hannah, A.F. Alrheli, D. Ancius et al., *Muon Imaging.* (International Atomic Energy Agency, Vienna, 2022), pp. 5-8

[7] H.K.M. Tanaka, C. Bozza, A. Bross et al., *Muography.* Nat. Rev. Methods Primers 3, 88 (2023). doi:10.1038/s43586-023

[8] L.W. Alvarez, J.A. Anderson, F.E. Bedwei et al., *Search for hidden chambers in the pyramids.* Science. 167, 3919 (1970). doi:10.1126/science.167.3919.832

[9] S. Chatzidakis, C.K. Choi, L.H. Tsoukalas, *Analysis of spent nuclear fuel imaging using multiple coulomb scattering of cosmic muons.* IEEE. T. Nucl. Sci. 63, 2866-2874 (2016). doi:10.1109/TNS.2016.2618009

[10] G. Leone, H.K.M. Tanaka, M. Holma et al., *Muography as a new complementary tool in monitoring volcanic hazard: implications for early warning systems.* Proc. R. Soc. A. 477, 2255 (2021). doi:10.1098/rspa.2021.0320

[11] R. Han, Q. Yu, Z. Li et al., *Cosmic muon flux measurement and tunnel overburden structure imaging.* JINST. 15, 06019 (2020). doi:10.1088/1748-0221/15/06/P06019

[12] Y.P. Cheng, R. Han, Z.W. Li et al., *Imaging internal density structure of the Laoheishan volcanic cone with cosmic ray muon radiography.* Nucl. Sci. Tech. 33, 88 (2022). doi:10.1007/s41365-022-01072-4

[13] K.N. Borozdin, G.E. Hogan, C. Morris et al., *Radiographic imaging with cosmic-ray muons.* Nature. 422, 277 (2003). doi:10.1038/422277a

[14] W.C. Priedhorsky, K.N. Borozdin, G.E. Hogan et al., *Detection of high-Z objects using multiple scattering of cosmic ray muons.* Rev. Sci. Instrum. 74, 4294-4297 (2003). doi:10.1063/1.1606536

[15] L.J. Schultz, K.N. Borozdin, J.J. Gomez et al., *Image reconstruction and material Z discrimination via cosmic ray muon radiography.* Nucl. Instrum. Meth. A. 519, 687-694 (2004). doi:10.1016/j.nima.2003.11.035

[16] P Lalor, A Danagoulian, *Fundamental limitations of dual energy X-ray scanners for cargo content atomic number discrimination.* Appl. Radiat Isotopes. 206, 111201 (2024). doi:10.1016/j.apradiso.2024.111201

[17] S. Xiao, W.B. He, M.C. Lan et al., *A modified multi-group model of angular and momentum distribution of cosmic ray muons for thickness measurement and material discrimination of slabs.* Nucl. Sci. Tech. 29, 28 (2018). doi:10.1007/s41365-018-0359-8

[18] X.T. Ji, S.Y. Luo, Y.H. Huang et al., *A novel 4D resolution imaging method for low and medium atomic number objects at the centimeter scale by coincidence detection technique of cosmic-ray muon and its secondary particles.* Nucl. Sci. Tech. 33, 2 (2022). doi:10.1007/s41365-022-00989-0

[19] P. Yu, Z.w. Pan, Z.Y. He et al., *A new efficient imaging reconstruction method for muon scattering tomography.* Nucl. Instrum. Meth. A. 1069, 169932 (2024). doi:10.1016/j.nima.2024.169932

[20] X.C. Ming, H.F. Zhang, R.R. Xu et al., *Nuclear mass based on the multi-task learning neural network method.* Nucl. Sci. Tech. 33, 48 (2022). doi:10.1007/s41365-022-01031-z

[21] G.L Wang, K.M. Chen, S.W. Wang et al., *Beam based alignment using a neural network.* Nucl. Sci. Tech. 35, 75 (2024). doi:10.1007/s41365-024-01436-y

[22] Y. Xie, Q. Li, *A review of deep learning methods reconstruction and its medical applications.* Electronics. 11(4), (2022). doi:10.3390/electronics11040586

[23] A. Bazzi, D.T.M. Slock, L. Meilhac et al., *A comparative study of sparse recovery and compressed sensing algorithms with application to AoA estimation.* Paper presented at IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Edinburgh, UK, July 2016. doi:10.1109/SPAWC.2016.7536780

[24] A. Bazzi, D.T. Slock, and L. Meilhac, *A Newton-type Forward Backward Greedy method for multi-snapshot compressed sensing.* Paper presented at 51st Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, Oct. 2017. doi:10.1109/ACSSC.2017.8335537

[25] A. Radovic, M. Williams, D. Rousseau et al., *Machine learning at the energy and intensity frontiers of particle physics.* Nature. 560, 41 (2018). doi:10.1038/s41586-018-0361-2

[26] G. Karagiorgi, G. Kasieczka, S. Kravitz et al., *Machine learning in the search for new fundamental physics.* Nat. Rev. Phys. 4, 399-412 (2022). doi:10.1038/s42254-022-00455-1

[27] A. Boehnlein, M. Diefenthaler, N. Sato et al., *Colloquium: Machine learning in nuclear physics.* Rev. Mod. Phys. 94, 031003 (2022). doi:10.1103/RevModPhys.94.031003

[28] C.Y. Gao, X.Z. Tang, X.N. Chen et al., *Convolutional neural network algorithm for material discrimination in muon scattering tomography.* Atom. Energy. Sci. Techno. 57, 353-361 (2023). doi:10.7538/yzk.2022.youxian.0055

[29] W.B. He, *Dissertation*, Nuclear Science and Technology University of Science and Technology of China, 2019 (in Chinese)

[30] W.B. He, D.Y. Chang, R.G. Shi et al., *Material discrimination using cosmic ray muon scattering tomography with an artificial neural network.* Radiat. Detect. Technol. Methods. 6, 254-261 (2022). doi:10.1007/s41605-022-00319-3

[31] S.J. Pan, Q. Yang, *A Survey on Transfer Learning.* IEEE. T. Knoml. Data. En. 22, 1345 (2010). doi:110.1109/TKDE.2009.191

[32] F. Zhuang, Z.Y. Qi, K.Y. Duan et al., *A Comprehensive Survey on Transfer Learning.* in Proceedings of the IEEE. 109, 43 (2021). doi:10.1109/JPROC.2020.3004555

[33] F.Z. Zhuang, P. Luo, Q. He et al., *Survey on Transfer Learning Research.* J. Softw. 26, 26-39 (2015). doi:10.13328/j.cnki.jos.004631

[34] L.J. Schultz, G.S. Blanpied, K.N. Borozdin et al., *Statistical Reconstruction for Cosmic Ray Muon Tomography.* IEEE. T. Image Process. 16, 1985-1993 (2007). doi:10.1109/TIP.2007.901239

[35] A Clarkson, D.G. Ireland, R.A. Jebali et al., *Characterising encapsulated nuclear waste using cosmic-ray Muon Tomography (MT).* Paper presented at the 4th International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA), Lisbon, Portugal, Apr. 2015. doi:10.1109/ANIMMA.2015.7465529

[36] J.M. Durham, *Cosmic Ray Muon Radiography Applications in Safeguards and Arms Control* (2018). arXiv:1808.06681

[37] J. Yosinski, J. Clune, Y. Bengio et al., *in Advances in Neural Information Processing Systems 27 (NeurIPS 2014)*, Montreal, Canada, December 2014 (Curran Associates, Inc., 2014), pp.

[38] Y. Ganin, E. Ustinova, H. Ajakan et al., *in Domain Adaptation in Computer Vision Applications*, ed. by G. Csurka (Springer, Cham, 2017), pp. 189-209. doi:10.1007/978-3-319-

[39] S. Agostinelli, J. Allison, K. Amako et al., *Geant4—a simulation*

toolkit. Nucl. Instrum. Meth. A. 506, 250-303 (2003). doi:10.1016/S0168-9002(03)01368-8

[40] C. Hagmann, D. Lange, and D. Wright, *Cosmic-ray shower generator for Monte Carlo transport codes.* Paper presented at the 2007 IEEE Nuclear Science Symposium Conference Record, Honolulu, HI, USA, Oct. 2007. doi:10.1109/NSSMIC.2007.443720

[41] A. Halevy, P. Norvig, F. Pereira, *The unreasonable effectiveness of data.* IEEE. INTELL. SYST. 24, 8-12 (2009). doi:10.1109/MIS.2009.36

[42] J. Deng, W. Dong, R. Socher et al., *ImageNet: A large-scale hierarchical image database.* Paper presented at the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, June 2009. doi:10.1109/CVPR.2009.5206848

[43] Y.C. Chen, *A tutorial on kernel density estimation and recent advances.* Biostatistics & Epidemiology, 1, 161-187 (2017). doi:10.1080/24709360.2017.1396742

[44] Goodfellow, Y. Bengio, A. Courville, *Deep Learning.* (MIT Press, Cambridge, 2016). pp. 131-134

[45] B. Sun, K. Saenko, *Deep CORAL: Correlation Alignment for Deep Domain Adaptation.* Paper presented at the ECCV 2016 Workshops, Amsterdam, The Netherlands, Oct. 2016. doi:10.1007/978-3-319-49409-8_{35}

[46] A. Gretton, K.M. Borgwardt, M.J. Rasch et al., *A Kernel Two-Sample Test.* J. MACH. LEARN. RES. 13, 723-773 (2012). http://jmlr.org/papers/v13/gretton12a.html

[47] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., *in Advances in Neural Information Processing Systems 27 (NeurIPS 2014)*, Montreal, Canada, December 2014 (Curran Associates, Inc., 2014), pp. 2672-2680.

[48] Y.P. Li, X.Z. Tang, X.N. Chen et al., *Experimental study on material discrimination based on muon discrete energy.* Acta Phys. Sin. 72, 026501 (2023). doi:10.7498/aps.72.20221645

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

Source: ChinaXiv — Machine translation. Verify with original.