Neutron and gamma discrimination based on multi-features and KNN-LDA

Authors: Yuhang, Mr. Jiang, Ding, Dr. Tingmeng, Ms. Xuanxi Wang, Jiang, Xiaofei, Miss Xiaofei Jiang

Date: 2025-03-02T00:00:00+00:00

Abstract

To improve n/γ discrimination performance, this paper proposes an intelligent discrimination method based on multi-features and the KNN-LDA algorithm. Firstly, this paper establishes six feature parameters covering both time and frequency domains based on the PSD principle; secondly, an automatic feature extraction system is designed, capable of intelligently extracting optimal distributions of multi-features from pulse data. Then, a feature criterion is constructed and calculated based on the feature distributions, to partition reliable data from the feature data as the model's training set, and the remaining data as the test set. Finally, using the training set, leveraging the KNN-LDA model's regression optimization and dimensionality reduction, the test set is fed into the model for classification to achieve n/γ discrimination for all pulses. Experimental results indicate that the KNN-LDA model achieves an FOM value of 3.07 in the high-energy domain (\$ $40keV) testset, representinga245$ \$40 keV) test set, the KNN-LDA model achieves an FOM value of 2.64, representing a 355% improvement over the CCM. These results demonstrate that the discrimination performance of this method is excellent, particularly for low-energy domain discrimination, which shows substantial improvement, providing a novel approach for introducing supervised learning models to address the challenging problem of low-energy domain discrimination.

Full Text

Preamble

Neutron and Gamma Discrimination Based on Multi-Features and KNN-LDA

Yuhang Jiang,[1] Tingmeng Ding,[1] Xuanxi Wang,[1] and Xiaofei Jiang[1],†
[1]Big Data and Information Engineering, Guizhou University, Guiyang 550025, China

To improve n/γ discrimination performance, this paper proposes an intelligent discrimination method based on multi-features and the KNN-LDA algorithm. First, we establish six feature parameters covering both time and frequency domains according to the PSD principle. Second, we design an automatic feature extraction system that intelligently extracts optimal distributions of multi-features from pulse data. Then, we construct and calculate a feature criterion to separate reliable data from the feature data as the training set for the model, with the remainder serving as the test set. Finally, based on the training set, we employ the KNN-LDA model for regression optimization and dimensionality reduction, input the test set into the model for further classification, and achieve n/γ discrimination for all pulses. Experimental results show that the FOM value of the KNN-LDA model reaches 3.07 on the high-energy domain ($\geq$ 40 keV) test set, which is 245% higher than that of CCM; on the low-energy domain ($\leq$ 40 keV) test set, the FOM value reaches 2.64, which is 355% higher than that of CCM. The results demonstrate that this method achieves excellent discrimination performance, particularly showing significant improvement in low-energy domain discrimination, providing a new approach for introducing supervised learning models to solve the challenging problem of low-energy domain discrimination.

Keywords: n/γ discrimination, multi-features, linear discriminant analysis (LDA), K-nearest neighbor (KNN), automatic feature extraction

Introduction

Discrimination between neutrons and γ-rays is a key issue in neutron detection research [?, ?, ?]. Discrimination methods based on the PSD (Pulse Shape Discrimination) principle have become the mainstream approach [?, ?, ?], and in recent years, with the rapid development of machine learning, intelligent discrimination methods based on machine learning have also emerged as a hot research direction in this field [?, ?, ?]. Traditional PSD-based discrimination methods typically use a single feature to visualize pulse shape differences [?, ?, ?], employing feature distributions to characterize the differentiation between neutrons and γ-rays to achieve n/γ discrimination. However, when the pulse shape difference between neutrons and γ-rays is small—particularly when the difference expressed by a given feature parameter is minor—the discrimination capability of that feature is greatly reduced. This situation frequently occurs in the low-energy domain [?, ?, ?] and represents the primary reason why pulse data discrimination is more challenging in the low-energy domain.

Intelligent discrimination methods based on machine learning have proven effective for enhancing low-energy domain discrimination performance. Among these, supervised learning algorithms offer advantages including strong interpretability, high prediction accuracy, and wide applicability, and have been applied multiple times in this field [?, ?, ?]. However, supervised learning algorithms require reliable training samples, yet no standard database of neutron and γ-ray pulses exists for model training in the n/γ discrimination domain.

This significantly weakens the portability of supervised learning models and substantially increases the difficulty of applying such models in field discrimination.

To address these challenges with difficult-to-discriminate pulse data and effectively improve n/γ discrimination performance while increasing model robustness, this paper proposes an intelligent discrimination method based on multi-features and KNN-LDA. First, we establish six waveform feature parameters to jointly characterize pulse differences from multiple dimensions and design an intelligent feature extraction algorithm to automatically obtain optimal distributions for each feature. Second, we synthesize the distribution of each feature to construct a feature criterion that classifies reliable and easily discriminable feature data as a training set, with the remaining data serving as the test set. Then, we train the KNN-LDA model using the training set, perform regression optimization and dimensionality reduction classification on the remaining test set, and finally achieve n/γ discrimination for all pulses.

Methods

A. Selected Multi-Features

To improve discrimination accuracy, this study proposes constructing feature parameters covering both time and frequency domains for n/γ discrimination. However, from the perspective of the discrimination model, the computational complexity of feature parameters must be considered. If the construction and computation of each feature parameter is cumbersome, the application of the entire multi-feature parameter set will face difficulties in feature extraction, which would defeat the purpose. Therefore, the selected feature parameters should express pulse differences across different dimensions while also being simple to construct and easy to extract, enabling efficient multi-feature extraction—a prerequisite for applying multidimensional feature parameters to n/γ discrimination.

Normalized neutron and γ-ray pulses typically differ in the falling edge segment [**?**, **?**, **?**], as shown in [Figure 1: see original paper]a. This difference can be concretized in the transverse direction, longitudinal direction, and integral domain. Accordingly, we construct two features in the transverse direction: pulse width (PW) and tail length (TL). Here, pulse width does not narrowly refer to the width at 50% amplitude but rather the width between the intersection points of the rising and falling edges at a certain amplitude level. Similarly, tail length refers to the width from a certain sample point on the falling edge to the last sample point of the pulse. In the longitudinal direction, we construct the tail height (TH) feature, representing the amplitude at a certain point along the falling edge. We also construct the tail integral (TI) feature, representing the integral value of a certain portion of the falling edge. These four time-domain features are shown in [Figure 2: see original paper]a and [Figure 2: see original paper]b.

Normalized neutron and γ-ray pulses also exhibit obvious differences in the low-frequency range of the amplitude spectrum after FFT [?, ?], as shown in [Figure 1: see original paper]b. Drawing on PSD features from traditional frequency-domain discrimination methods, we construct zero-frequency amplitude (ZF-AMP) and frequency gradient (FG), which have been utilized effectively for n-γ discrimination in previous work [?, ?], demonstrating good discrimination capability, as shown in [Figure 2: see original paper]c.

This study combines traditional discrimination methods, selecting feature parameters from multiple viewpoints in the time-frequency domain to comprehensively characterize the difference information contained in pulse waveforms. By applying these multidimensional feature parameters to n-γ discrimination research, we aim to obtain better discrimination results, particularly achieving improved discrimination performance and more reliable classification results when pulse waveform differences are smaller.

B. Automatic Feature Extraction

In PSD-based discrimination methods [?, ?, ?], the distribution of different PSD factors is generally obtained by continuously adjusting feature measurement parameters, from which the optimal distribution is determined. The valley between the two distribution peaks is then used as the discrimination threshold to achieve n/γ discrimination. However, the process of finding optimal feature distributions requires manual adjustment, and the optimal parameters differ under various experimental scenarios [?, ?, ?], making this process cumbersome and repetitive. Moreover, due to human judgment, the best results may not ultimately be obtained. Therefore, achieving automation of feature extraction is crucial [?].

The key to automating feature extraction is finding and outputting the feature distribution that offers the best separation, which requires comparing the separation degree of feature distributions obtained from different measurement parameters. Traditional methods use FOM (Figure Of Merit) to evaluate n/γ discrimination effect [?], expressing the degree of separation between the neutron peak and γ-ray peak. For distinction, this paper uses DOS (Degree Of Separation), defined as shown in [Figure 3: see original paper], where d is the spacing between the two peaks, and $FWHM_n$ and $FWHM_\gamma$ are the full width at half maximum of the neutron peak and γ-ray peak, respectively.

The DOS value can be calculated by Eq. (1), where a larger DOS value indicates greater separation. By calculating and comparing DOS values of feature distributions measured with different parameters, we can find the optimal feature distribution needed, thus achieving automated feature extraction.

The automatic extraction process for a single feature is shown in [Figure 4: see original paper]. First, based on the input preprocessed pulse and the feature to be measured, the range of extraction parameters for the feature is confirmed. Second, the system enters a loop structure where a set of measurement parameters is input into each iteration, and the program measures the features of all pulses according to these parameters, obtaining a feature distribution. Then, the DOS values of the neutron peak and γ-ray peak are calculated and compared with the DOS value stored in the previous loop in the shift register, retaining the larger value along with its corresponding feature measurement value. Finally, when the loop ends, the largest value and its corresponding feature measurement value are output, thereby accomplishing automatic feature extraction while ensuring the best degree of separation.

In this process, the most critical aspect is how the program autonomously calculates DOS. Due to the indeterminate shape of feature distribution curves and the inability to directly obtain DOS calculation parameters, we uniformly perform Gaussian fitting on the distribution curves and use the fitting parameters to directly calculate DOS, as in Eq. (2), where μ_n, μ_γ and σ_n, σ_γ are the mean and standard deviation of the neutron and γ-ray fitting peaks, respectively. The mean and standard deviation can be obtained directly after performing Gaussian fitting.

However, for feature distribution curves, the program cannot perform Gaussian fitting of bimodal peaks by itself. Therefore, we switch the approach to split the bimodal distribution from the troughs between them, forming two single peaks that can be automatically Gaussian-fitted simultaneously. This converts the problem into how to correctly identify the fitting trough.

Because feature distribution curves are affected by many factors—such as the size of the pulse dataset, feature category, and measurement location of the same feature—the distribution curve shape is not necessarily a regular bimodal shape and is more likely to have various burrs and fluctuations. This makes it difficult to determine the location of the target trough due to the presence of numerous meaningless troughs during trough searching. Therefore, before seeking the trough, this paper performs three-time spline fitting (with balance parameter set to 0.9) to smooth the distribution curve without deformation. Then, based on the position of the double peaks, we use the determined valley to split the original distribution curve before fitting, perform Gaussian fitting simultaneously, and then calculate the DOS value.

In this way, the program can autonomously find the best feature distribution, optimizing the manual parameter tuning process.

C. Feature Criterion

After obtaining feature data for pulses, each feature is classified according to its distribution. This paper constructs Comprehensive Voting Factor (CVF) and Comprehensive Location Factor (CLF) to comprehensively evaluate each pulse's classification performance across multi-features. Based on these metrics, reliable datasets are separated from feature data to serve as training samples, with the remaining data used as relatively more difficult test samples. The trained KNN-LDA model then further discriminates the test set.

Initial discrimination through feature distributions from automatic feature extraction (marked as 1 for neutron, 0 for γ-ray) yields six discrimination result datasets, from which we construct an $M \times 6$ matrix of 0s and 1s to represent these results, as in Eq. (3). The computed weights W of CVF are calculated by Eq. (4). We obtain a matrix $B_{6\times 1}$ as in Eq. (5), and CVF can be calculated as in Eq. (6).

The values $CVF(i)$ in the $CVT_{M\times 1}$ matrix range from 0 to 1. We use Eq. (7) to determine the category of the ith pulse, realizing the comprehensive voting result with multi-features. Obviously, if $CVF(i) = 0$, all six features classify the pulse as a γ-ray pulse; if $CVF(i) = 1$, all six features classify it as a neutron pulse. For pulses with CVF values of 0 and 1, we consider their discrimination results highly reliable and suitable for constructing the training set. The closer CVF is to 0.5, the more disagreement occurs in multi-feature discrimination, and the less reliable the results become.

CVF integrates the category judgment of each feature for each pulse. For feature distributions output from the automatic feature extraction system, each pulse may perform differently across features—that is, a pulse may be far from the trough in one feature distribution but immediately adjacent to it in another. Therefore, a single feature may not fully represent pulse shape differences. Moreover, higher quality training samples lead to better supervised model performance. Consequently, we construct CLF to further optimize the reliable data selected by CVF based on relative positions in feature distributions, thereby obtaining a high-quality training set.

As shown in [Figure 5: see original paper], which illustrates a typical PSD feature histogram distribution curve (assuming the left peak is the γ-peak and the right peak is the neutron peak), the PSD feature value for the ith pulse is $c(i)$, and the trough value is v. Based on this, we define the relative position of each pulse on this feature using Eq. (8), where $L(i)_{feature}$ refers to the relative position of the ith pulse on the feature; a and b are the distances from the gamma and neutron peaks to the trough, respectively; and M represents the number of mixed pulses.

Obviously, the relative position of a pulse with a feature value at the γ-peak is negative, at the neutron peak is non-negative, and at the center of the two peaks is exactly ± 1. The larger the absolute value of the relative position and the further from the zero point, the better the separation at that feature. The distribution of relative positions for a single feature is fundamentally similar to the shape of the original distribution, providing the conditions to synthesize feature distributions for each pulse. We define CLF directly in Eq. (9).

It can be seen that the larger the absolute value of CLF, the farther the pulse's relative position is from the classification boundary point, the better its combined performance across feature distributions, and the higher the corresponding classification reliability.

In summary, we use CVF and CLF together as constraints to form the feature criterion in Eq. (10). CVF constrains multi-feature category reliability, while CLF constrains multi-feature positional reliability, enabling selection of a high-quality feature training set.

D. KNN-LDA

K-Nearest Neighbor (KNN) is a commonly used supervised learning method [?]. Its working mechanism is straightforward: given a test sample, find the K closest samples in the training set based on a certain distance metric, then make predictions based on information from these K "neighbors." KNN is an instance-based learning model where the training phase only stores training samples, and distances are calculated upon receiving test samples. Because each test sample must calculate distances to all training samples, the testing time overhead is substantial. Therefore, this paper uses the six extracted features as KNN inputs to reduce computational complexity while ensuring pulse difference information is fully retained.

In this paper, the distance between test and training samples is calculated using Euclidean distance as in Eq. (11), where $D(X, Y)$ is the distance between samples X and Y, m is the number of features contained in the samples, and x_i and y_i are the corresponding ith feature values in the two samples.

After distance calculation and comparison, the K training samples closest to the test sample are obtained, and the average of these K samples is used as the regression result via the "averaging method." Based on the training set of feature data, the test set can be regressed by KNN, and the multi-dimensional characteristics of feature data support the reasonableness of regression features used for classification, enabling these regression features to serve as LDA input for further dimensionality reduction and classification.

Linear Discriminant Analysis (LDA) [?] is a classical linear learning method first proposed by Fisher in 1936 for binary classification problems. Moreover, LDA's projection process achieves dimensionality reduction while retaining category information, making it a classical supervised dimensionality reduction technique. The idea of LDA [?] is to project a given set of training samples onto a straight line such that projection points of the same class are as close as possible while projection points of different classes are as far apart as possible. When classifying a new sample, it is projected onto this line, and the sample's class is determined based on its projection point position, as shown in [Figure 6: see original paper].

Pulse data can be represented by the matrix in Eq. (12), which contains a total of m digitized pulses of neutrons mixed with γ-rays, each discretized into l sampling points. Let $x_n = (x_{n1}, x_{n2}, \ldots, x_{nl})$ denote neutron pulses and $x_\gamma = (x_{\gamma 1}, x_{\gamma 2}, \ldots, x_{\gamma l})$ denote γ-ray pulses.

Let the mean vectors and covariance matrices of x_n and x_γ be μ_n, μ_γ and Σ_n, Σ_γ respectively. If the pulse data are projected onto a straight line w, the projections of the centers of the two classes onto line w will be $w^T\mu_n$ and $w^T\mu_\gamma$, and the covariances of the two classes will be $w^T\Sigma_n w$ and $w^T\Sigma_\gamma w$ respectively. Our goal is to maximize the inter-class distance between the two types of projection points while minimizing their intra-class distance—that is, to make the center distance $||w^T\mu_n - w^T\mu_\gamma||^2$ as large as possible and the sum of covariances $w^T\Sigma_n w + w^T\Sigma_\gamma w$ as small as possible. Considering both simultaneously yields the maximization objective in Eq. (13).

Let the within-class scatter matrix be $S_w = \Sigma_n + \Sigma_\gamma$ and the between-class scatter matrix be $S_b = (\mu_n - \mu_\gamma)(\mu_n - \mu_\gamma)^T$. Substituting into Eq. (13) yields Eq. (14). Through derivation, we can calculate that when line w satisfies Eq. (15), J obtains a maximum value. Considering numerical solution stability, S_w^{-1} is usually obtained through singular value decomposition of S_w, i.e., $S_w = U\Sigma V^T$, where Σ is a real diagonal matrix whose diagonal elements are the singular values of S_w. S_w^{-1} can then be obtained from Eq. (16), from which the optimal projection direction w can be determined.

From the above principle analysis, we see that the KNN regression algorithm and LDA classification algorithm are simple in principle, easy to calculate, and straightforward to apply. We can make full use of and combine their regression optimization and dimensionality reduction classification functions. In summary, the KNN-LDA model obtains KNN regression of the test set through training set data, acquires the optimal LDA projection vector, and then uses this vector to project regression features down to a one-dimensional space to obtain discrimination results for one-dimensional features.

Experiment

LabVIEW is a graphical development environment commonly used to build digital n-γ discrimination platform programming software. It can construct user interactive interfaces to facilitate signal display and parameter modification. Moreover, its graphical programming is simple and easy to learn, while also embedding various directly usable data processing modules that greatly reduce programming difficulty. It is widely used in industrial automation, test and measurement, signal processing, and other fields [**?**, **?**, **?**]. As shown in [Figure 7: see original paper], this paper constructs the flowchart of the multi-features based KNN-LDA model discrimination using LabVIEW.

First, preprocessed pulse data is input, from which the automatic feature extraction algorithm extracts multi-features. Then, corresponding CVF and CLF are calculated based on the output multi-features classification results and distribution locations, selecting high-quality training samples from the feature data to complete initial classification. Finally, the training samples are delivered to the KNN-LDA model, where the model is trained to perform regression and dimensionality reduction classification on the remaining fuzzy samples (test set), and classification results are output, thus completing the discrimination task.

A. Data Acquisition and Processing

The experimental dataset in this paper contains 40,000 pulse signals obtained by an EJ-301 detector detecting a ^{241}Am-Be neutron source. The neutron detection experimental environment and data acquisition schematic are shown in [Figure 8: see original paper]. During the experiment, the detector is placed horizontally at a 0-degree angle relative to the radioactive source, with the distance between the neutron source and detector set to 40 cm. The detected pulse signals are processed by a digitizer at 500 MS/s and transmitted to a PC for preservation. Each pulse waveform contains 128 sampling points, with a 2 ns time interval between neighboring sampling points.

In the process of converting neutrons and γ-rays into pulse signals through detectors, electronic noise and baseline drift occur, necessitating pulse preprocessing to eliminate these effects and unify the premise for feature extraction. For the pulse dataset in this paper, we primarily perform baseline zeroing, flipping, smoothing filtering, and normalization operations.

We select the 30 sample points at the end of the pulse waveform, obtain their mean value, and subtract this mean from the entire pulse to complete simple baseline zeroing. Afterward, we flip the pulse by adding a minus sign to the amplitude of each sample point. Then, we apply a seven-point moving average smoothing filter to the signal, which essentially takes a total of 7 points (including 3 points before and after the current point plus the point itself), calculates their amplitude mean, and replaces the current point. Finally, we map the pulse between 0 and 1 to complete normalization. The preprocessing effect is shown in [Figure 9: see original paper].

Pulses with small waveform differences that are difficult to discriminate are mainly concentrated in the low-energy domain. To explore the ability of the proposed method to improve discrimination performance for this portion, we divide the preprocessed pulse dataset into a low-energy domain pulse dataset (containing 21,464 pulses, as dataset 1) and a high-energy domain pulse dataset (containing 18,536 pulses, as dataset 2) using the 40 keV threshold, and perform discrimination experiments on each dataset respectively.

B. Features Extraction

After inputting preprocessed pulses into the automatic feature extraction system, the system automatically obtains the optimal distribution for each feature and relies on trough positions to automatically perform single-feature discrimination. The results are shown in , with corresponding feature distributions and DOS values displayed in [Figure 10: see original paper], [Figure 11: see original paper], and [Figure 12: see original paper].

As shown in , taking neutron counts as an example, in the high-energy domain, the largest difference in particle counts discriminated by multi-features occurs between TH and TL, with a difference of 226 particles accounting for 1.2% of the total. In the low-energy domain, the largest difference occurs between TH and TI, with a difference of 1,747 particles accounting for 8.1% of the total—approximately 6.8 times that of the high-energy domain. Alternatively, if we calculate the mean difference in discriminated particle counts between any two different features, we obtain a mean difference of about 91 (0.49% of particles) in the high-energy domain and about 786 (3.7% of particles) in the low-energy domain—approximately 7.6 times larger. These analyses demonstrate that discrimination in the low-energy domain involves greater uncertainty than in the high-energy domain, making low-energy domain discrimination more difficult. This also directly shows that the reliability of a single feature parameter decreases dramatically in the low-energy domain compared to the high-energy domain. Therefore, the comprehensive use of multidimensional feature parameters for n/γ discrimination is undoubtedly a direct and effective way to improve discrimination reliability in the low-energy domain.

As evident from the feature distributions and [Figure 12: see original paper], the DOS of each time-domain feature distribution in the low-energy domain is significantly smaller than in the high-energy domain, with much better bimodal separation in the high-energy domain, clearly indicating that high-energy domain data are more easily discriminated. Moreover, TI performs poorly in the low-energy domain but well in the high-energy domain, where the DOS value changes most dramatically, suggesting it may be more suitable for high-energy domain discrimination tasks. Conversely, the two frequency-domain features show good DOS performance in both low- and high-energy domains, suggesting that frequency-domain discrepancy provides better discrimination capability and stronger stability. Therefore, discriminating pulse data with multi-features covering the time-frequency domain is an intuitive and effective approach to improve discrimination results.

When multi-features are output, the program calculates the CVF and CLF for each pulse based on the classification results and distribution locations of each feature. The resulting distribution histograms are shown in [Figure 13: see original paper] and [Figure 14: see original paper].

These figures show the comprehensive classification results of multi-features in terms of categories and distributions. From the CVF histogram, pulses with CVF values of 0 and 1 constitute the main components of the pulse dataset, representing consistent classification voting results across multi-features. Values between 0 and 1 represent contradictory classification voting results, with CVF values closer to 0.5 indicating deeper contradictions. Compared to the high-energy domain, the distribution between 0 and 1 in the low-energy domain plot shows obvious pile-up. Calculations reveal that this portion accounts for 12.23% of pulses in the low-energy domain dataset but only 2.42% in the high-energy domain dataset—approximately 10% less—clearly verifying that pulse data classification is more difficult in the low-energy domain. The CLF histogram forms an obvious double-peak shape with zero as the trough and ± 1 as the peaks, with more serious pile-up at zero in the low-energy domain and obviously better double-peak separation in the high-energy domain, more intuitively showing that waveform differences among low-energy domain pulses are even more minute.

To improve discrimination performance on harder-to-discriminate pulse data, we introduce the supervised discrimination model KNN-LDA and use the feature criterion composed of CVF and CLF to instantly select high-quality training samples. According to Section 2.3, we select a total of 18,071 feature data sets (including 6,396 neutron feature data and 11,675 γ-ray feature data) from the low-energy domain dataset as the training set, with the remaining 3,393 feature data sets used as test set 1 for further discrimination. From the high-energy domain dataset, we select 16,931 feature data sets (including 8,564 neutron feature data and 8,367 γ-ray feature data) as the training set, with the remaining 1,605 feature data sets used as test set 2 for further discrimination. Subsequent discrimination experiments will focus on these residual test sets to compare model discrimination performance.

C. CCM and LDA

For comparison with KNN-LDA, this study selects two methods: the classical charge comparison method (CCM) and the LDA algorithm. CCM uses the integral ratio of different pulse signal intervals to discriminate particle types. Since CCM more fully utilizes shape differences between neutron and γ-ray pulses, it achieves good discrimination performance under most conditions and is widely used by researchers. This paper utilizes the ratio of short-gate integral Q_S to long-gate integral Q_L as the CCM discrimination factor, as in Eq. (17).

First, we validate the training set data using CCM. As shown in [Figure 15: see original paper], the CCM histograms of the two training sets achieve FOM values of 1.20 in the low-energy domain and 1.15 in the high-energy domain, both demonstrating good separation. Using the trough as the discrimination threshold, the classification results for both datasets are fully consistent with the feature criterion labeling, demonstrating that the training set data instantly divided by the feature criterion is highly reliable and suitable for model training.

Next, we apply CCM discrimination to the two test sets, with resulting distribution histograms shown in [Figure 16: see original paper]. The bimodal separation from the low-energy domain test set is very poor, with a FOM of only 0.58, while the high-energy domain is relatively better but still only achieves a FOM of 0.89. Obviously, achieving satisfactory discrimination results with CCM for difficult-to-differentiate impulse data is very challenging, especially in the low-energy domain.

Finally, we input the training set feature data into the LDA model for training and reduce the test set feature data to one-dimensional feature data according to the obtained optimal projection direction. The resulting distribution histogram is shown in [Figure 17: see original paper]. The LDA model's discrimination effect is better than CCM, but the FOM in the low-energy domain only improves by 0.08 to 0.65, while in the high-energy domain the FOM improves by 0.23 to 1.12. The LDA model's discrimination improvement is more obvious for the high-energy domain portion but remains insufficient for the low-energy domain portion.

D. KNN-LDA

The KNN-LDA algorithm requires setting a suitable value for the nearest neighbor K, which significantly impacts algorithm performance and computational cost. Considering computational efficiency and the high-density distribution characteristics of the feature dataset, setting a smaller K value satisfies algorithm requirements. We set the K value range to [3,5,7] and verify that KNN achieves the best classification accuracy when $K = 5$ using the labeled training set. Therefore, in this paper, K is set to 5, meaning the mean value of the five training samples closest to the input sample is used as the regression result for that input sample in KNN regression.

Afterward, the training set data is input into the KNN-LDA model for training, after which the test set data is regressed and downscaled. The one-dimensional feature distribution histogram output from the model is shown in [Figure 18: see original paper]. As seen in the figure, the KNN-LDA model demonstrates excellent discrimination performance, with clearly distinguished double peaks in both low-energy and high-energy domains. The FOM reaches 2.64 in the low-energy domain and 3.07 in the high-energy domain, representing a significant improvement in discrimination performance.

E. Results and Analysis

The FOM indicators obtained from each method are summarized in .

As shown in , KNN-LDA demonstrates brilliant FOM performance. In the high-energy domain test set, the FOM obtained from the KNN-LDA model is 2.18 higher than CCM—an enhancement of about 245%—and 1.95 higher than LDA—an enhancement of about 174%. In the low-energy domain test set, the FOM is 2.06 higher than CCM—an enhancement of about 355%—and 1.99 higher than LDA—an enhancement of about 306%. This shows that KNN regression can effectively optimize feature data in the multidimensional feature space, enabling the KNN-LDA model to significantly improve discrimination performance, particularly showing very powerful enhancement for the low-energy domain portion.

Additionally, comparing FOM values across different energy domains for the same method reveals that high-energy domains are significantly better than low-energy domains, again verifying that high-energy domain pulse data have more obvious waveform differences and are easier to discriminate, even though they all belong to the more difficult-to-discriminate portion of the original dataset.

The experimental results can be analyzed as follows: First, the proposed multi-features-based KNN-LDA model shows better discrimination performance compared to traditional single-feature-based methods, particularly for low-energy domain data, demonstrating the advantage of multi-feature comprehensiveness and supporting the reasonableness of the KNN-LDA model. Second, the KNN-LDA model effectively mines and optimizes the discrimination capability of multi-features. The strategy of constructing a training set for the supervised model from multi-features in the original dataset and then performing high-precision classification using the model is both effective and feasible. This not only facilitates field discrimination application of the supervised model but also makes algorithm transplantation feasible to some extent. Lastly, the automatic feature extraction algorithm designed in this paper can effectively obtain optimal feature distributions, optimizing the feature extraction process and greatly enhancing the feasibility of this method.

Conclusion

This paper proposes an n/γ discrimination model based on multi-features and KNN-LDA. The model achieves integrated automation, where feature extraction, dataset splitting, model training, and test set discrimination are performed continuously and automatically, offering a complete system with diverse functionalities. Experimental results show that the model achieves a FOM of 3.07 on the high-energy domain test dataset and 2.64 on the low-energy domain test dataset, significantly improving discrimination performance. The model effectively discriminates low-energy domain pulses that are difficult for CCM to distinguish, providing a new approach for introducing supervised learning algorithms in the n/γ discrimination field to solve the challenging problem of low-energy domain discrimination and making transplantation application of supervised discrimination models more feasible.

References

[1] D.Z. Ding, C.T. Ye, Z.X. Zhao, et al., in *Neutron Physics: Principles, Methods, and Applications*, 2nd edn. (Atomic Energy Press, Beijing, 2005), p. 105

[2] A. Yamazaki, K. Watanabe, A. Uritani, et al., Neutron-gamma discrimination based on pulse shape discrimination in a Ce:LiCaAlF6 scintillator. *NUCL. INSTRUM. METH. A.* 652, 635-438(2011). DOI: 10.1016/j.nima.2011.02.064

[3] J.X. Li, H.L. Hou, Y.F. Huang, et al., Pulse-shaping method for real-time neutron/gamma discrimination at low sampling rates. *Nucl. Sci. Tech.* 34, 165(2023). DOI: 10.1007/s41365-023-01306-z

[4] Bo. W, Gang. L, Kun. L, et al., Digital pulse shape discrimination technique for liquid scintillation detector EJ-301. *JINST.* 14, T11001(2019). DOI: 10.1088/1748-0221/14/11/T11001

[5] J. Choi, J. Park, J. Son, et al., Study of n/γ discrimination using 3He proportional chamber in high gamma-ray fields. *NUCL. ENG. TECHNOL.* 51,

263-268(2019). DOI: 10.1016/j.net.2018.08.013

[6] D. Zhao, S. Feng, C. Hu, et al., Characterization of the neutron/γ-ray discrimination performance in an EJ-301 liquid scintillator for application to prompt fission neutron spectrum measurements at CSNS. *RADIAT. MEAS.* 151, 106703(2022). DOI: 10.1016/j.radmeas.2022.106703

[7] F. Wang, M. Yang, J. Wang, et al., A comparison of small-batch clustering and charge-comparison methods for n/γ discrimination using a liquid scintillation detector. *RADIAT. MEAS.* 1028, 166379(2022). DOI: 10.1016/j.nima.2022.166379

[8] L. Liu, H. Shao, Study on neutron–gamma discrimination method based on the KPCA-GMM. *NUCL. INSTRUM. METH. A.* 1065, 168604(2023). DOI: 10.1016/j.nima.2023.168604

[9] W. Hu, G. Zhang, Y. Zhang, et al., Neutron/gamma (n/γ) discrimination method based on KPCA-MPA-ELM. *Nucl. Tech.* 75-84(2024). DOI:10.11889/j.0253-3219.2024.hjs.47.040403

[10] J. Yan, R. Liu, C. Li, et al., A comparison of n/γ discrimination by the rise-time and zero-crossing methods. *SCI. CHINA PHYS. MECH.* 53, 1453-1459(2010). DOI: 10.1007/s11433-

[11] T. Szczcsniak, M. Grodzicka, M. Moszynski, et al., Digital Neutron-Gamma Discrimination Methods: Charge Comparison versus Zero-Crossing, in *IEEE Nuclear Science Symposium and Medical Imaging Conference Record*, pp.1-4(2014). DOI: 10.1109/NSSMIC.2014.7431222

[12] M. Nakhostin, A comparison of digital zero-crossing and charge-comparison methods for neutron/γ-ray discrimination with liquid scintillation detectors. *NUCL. INSTRUM. METH. A.* 797, 77-82(2015). DOI: 10.1016/j.nima.2015.06.041

[13] T. Ding, Y. Jiang, X.Wang, et al., Study on neutron-gamma discrimination methods based on GMM-KNN and LabVIEW implementation. *Nucl. Sci. Tech.* 35, 84-99(2024). DOI: 10.1007/s41365-024-01545-8

[14] S. Yoon, C. Lee, H. Seo, et al., Improved fast neutron detection using CNN-based pulse shape discrimination. *NUCL. ENG. TECHNOL.* 55, 3925-3934(2023). DOI: 10.1016/j.net.2023.07.007

[15] S. Zhang, Z.Wei, P. Zhang, et al., Neutron-gamma discrimination with broaden the lower limit of energy threshold using BP neural network. *APPL. RADIAT. ISOTOPES.* 205, 111179(2024). DOI: 10.1016/j.apradiso.2024.111179

[16] X. Yu, J. Zhu, S. Lin, et al. Neutron–gamma discrimination based on the support vector machine method. *NUCL. INSTRUM. METH. A.* 777, 80-84(2015). DOI: 10.1016/j.nima.2014.12.087

[17] H. Arahmane, A. Mahmoudi, E.-M. Hamzaoui, et al., Neutron-gamma discrimination based on SVM combined on nonnegative matrix factorization and continuous wavlet transform. *MEASUREMENT.* 149, 106958(2020). DOI: 10.1016/j.measurement.2019.106958

[18] H. Arahmane, E.M. Hamzaoui, Y. Ben Maissa, et al., Neutron-gamma discrimination method based on blind source separation and machine learning. *Nucl. Sci. Tech.* 32, 18(2021). DOI: 10.1007/s41365-021-00850-w

[19] S. Yousefi, L. Lucchese, M.D. Aspinall, Digital discrimination of neutrons and gamma-rays in liquid scintillators using wavelets. *NUCL. INSTRUM. METH. A.* 598, 549-553(2009). DOI: 10.1016/j.nima.2008.09.028

[20] K.N. Li, X.P. Zhang, Y. Li, et al., Investigation on n/γ Discrimination Methods for Liquid Scintillator Detector. *At. Energy Sci. Technol.* 5, 913-919(2014). DOI: 10.7538/yzk.2014.48.05.0913

[21] R.C. Pereira, A. Fernandes, N. Cruz, et al., Neutron/Gamma discrimination code based on trapezoidal filter. *FUSION. ENG. DES.* 134, 118-122(2018). DOI: 10.1016/j.fusengdes.2018.07.002

[22] G.F. Liu, M.J. Joyce, X.D. Ma, et al., A Digital Method for the Discrimination of Neutrons and γ Rays With Organic Scintillation Detectors Using Frequency Gradient Analysis. *IEEE T. NUCL. SCI.* 57, 1682-1691(2010). DOI: 10.1109/TNS.2010.2044246

[23] G. Liu, J. Yang, X.L. Luo, et al., A comparison of different discrimination parameters for the DFT-based PSD method in fast scintillators. *RADIAT. MEAS.* 58, 12-17(2013). DOI: 10.1016/j.radmeas.2013.07.008

[24] B. D'Mellow, M.D. Aspinall, R.O. Mackin, et al., Digital discrimination of neutrons and γ-rays in liquid scintillators using pulse gradient analysis. *NUCL. INSTRUM. METH. A.* 578, 191-197(2007). DOI: 10.1016/j.nima.2007.04.174

[25] J. Cai, D. Li, P. Wang, et al., Fast pulse sampling module for real-time neutron–gamma discrimination. *Nucl. Sci. Tech.* 30, 84(2019). DOI: 10.1007/s41365-019-0595-1

[26] S. Peng, Y. Yi, Y. Lu, et al., Neutron/Gamma Real-time Discrimination Based on Frequency Comparison Analysis. *At. Energy Sci. Technol.* 30, 520-525(2020). DOI: 10.7538/yzk.2018.youxian.0622

[27] S. Liu, W. Zhang, Z. Zhang, et al., Performance of real-time neutron/gamma discrimination methods. *Nucl. Sci. Tech.* 34, 8(2023). DOI: 10.1007/s41365-022-01160-5

[28] L.F. Liu, Research on n/γ discrimination method of EJ309 liquid scintillator detector. Master Dissertation. Beijing: Academy of Military Sciences, DOI:10.27193/d.cnki.gjsky.2022.000054

[29] P.V. Chuan, L.V. Tung, N.N. Hai, et al., A Study on the Impact of Pulse Shaping Parameters on Zero-Crossing Method Performance for Neu-

tron/Gamma Discrimination. *IEEE T. NUCL. SCI.* 70, 2464-2470(2023). DOI: 10.1109/TNS.2023.3323369

[30] Y. Jiang, T. Ding, X. Jiang, Automatic n-γ discrimination design based on feature intelligent extraction. *Nuclear Electronics and Detection Technology.* 44, 1091-1100(2024). DOI:10.20173/j.cnki.ned.20240903.001

[31] W. Bo, L. Gang, L. Kun, et al., Digital pulse shape discrimination technique for liquid scintillation detector EJ-301. *J. INSTRUM.* 14, T11001(2019). DOI:10.1088/1748-0221/14/11/T11001

[32] X. Wang, Q. Tang, X. Jiang, Simulation of n-γ pulse signal discrimination based on KNN classification algorithm. *Electronic Measurement Technology.* 45, 164-170(2022). DOI: 10.19651/j.cnki.emt.2209025

[33] K. Li, X. Zhang, M. Zhang, et al., n/γ Pulse Shape Discrimination in Cs2LiYCl6:Ce3+ Crystal Using Fisher Linear Discriminant. *At. Energy Sci. Technol.* 51, 2069-2074(2017). DOI: 10.7538/yzk.2017.51.11.2069

[34] Z.H. Zhou, in *Machine learning*, ed. by H. Xue (Peking University Press, Beijing, 2016), pp. 60-63

[35] M.E. Iranian, M. Mohseni, S. Aghili, et al., Real-Time FPGA-Based HIL Emulator of Power Electronics Controllers Using NI PXI for DFIG Studies. *IEEE J. EM. SEL. TOP. P.* 10, 2005-2019(2022). DOI: 10.1109/JESTPE.2020.3023100

[36] G. Dhanabalan, S.T. Selvi, Software Design of VerilogHDL Code Generation for Ladder Diagram and Data Acquisition Using LABVIEW. *WIRELESS PERS. COMMUN.* 128, 1087-1115(2023). DOI: 10.1007/s11277-022-09990-7

[37] R. Swarnalatha, S. Athiban, Analysis of brain wave data to detect epileptic activity using LabVIEW. *SOFT COMPUT.* 27, 17231–17241(2023). DOI: 10.1007/s00500-023-08047-6

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