KAN-ResCNN: An Advanced Deep Learning Model for Distinguishing Z Boson Decay Modes

Authors: Liu, Prof. Wufeng, Wang, Mr. Feihu, Jiang, Mr. Pengsong, Wei, Mr. Renjie, Zhang, Mr. Zhouli, Mr. Yuhong Yu, Mr. Xiangman Liu, Li, Mr. Yan, Li, Mr. Longfei, Liu, Prof. Wufeng

Date: 2025-02-24T00:00:00+00:00

Abstract

Addressing the growing demand for intelligent automation in particle physics analysis, this study introduces an innovative deep learning framework for classifying Z boson leptonic decay channels ($Z\rightarrow \mu^+\mu^-$ vs. $Z\rightarrow e^+e^-$). We propose KAN-ResCNN, a hybrid architecture synergistically integrating three advanced paradigms: Kolmogorov-Arnold Networks, Residual Networks, and Convolutional Neural Networks. The model employs a dual-branch feature extraction system where the Residual Networks backbone captures hierarchical physical patterns through residual learning while the Convolutional Neural Networks stream processes localized feature correlations. Theoretical physics principles are systematically encoded through Kolmogorov-Arnold Networks driven nonlinear classification, ensuring physically interpretable decision boundaries. Evaluated on a publicly available experimental dataset from CERN' s collider experiments, the proposed model achieves state-of-the-art performance with 94.73% classification accuracy and 94.75% F1-score, demonstrating statistically significant improvements over conventional machine learning baselines and monolithic deep architectures. This work establishes a novel paradigm for high-energy physics data analysis that simultaneously optimizes discriminative power and physical consistency. The architecture' s modular design permits direct extension to other quantum chromodynamics processes, providing critical technical infrastructure for next-generation intelligent particle physics analysis systems. Key implementation details and ablation studies are presented to elucidate the relative contributions of each architectural component.

Full Text

Preamble

KAN-ResCNN: An Advanced Deep Learning Model for Distinguishing Z Boson Decay Modes

Wufeng Liu,[1]*,† Feihu Wang,*[1] *Pengsong Jiang,*[1] *Renjie Wei,*[1] *Zhouli Zhang,*[2] *Yuhong Yu,*[2] *Xiangman Liu,*[2] *Yan Li,*[1] *and Longfei Li*[1]

[1]Henan University of Technology, Zhengzhou 450001, China

[2]Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou 730000, China

Addressing the growing demand for intelligent automation in particle physics analysis, this study introduces an innovative deep learning framework for classifying Z boson leptonic decay channels ($Z\to\mu^+\mu^-$ vs. $Z\to e^+e^-$). We propose KAN-ResCNN, a hybrid architecture that synergistically integrates three advanced paradigms: Kolmogorov-Arnold Networks, Residual Networks, and Convolutional Neural Networks. The model employs a dual-branch feature extraction system where the Residual Networks backbone captures hierarchical physical patterns through residual learning while the Convolutional Neural Networks stream processes localized feature correlations. Theoretical physics principles are systematically encoded through Kolmogorov-Arnold Networks-driven nonlinear classification, ensuring physically interpretable decision boundaries. Evaluated on a publicly available experimental dataset from CERN' s collider experiments, the proposed model achieves state-of-the-art performance with 94.73% classification accuracy and 94.75% F1-score, demonstrating statistically significant improvements over conventional machine learning baselines and monolithic deep architectures. This work establishes a novel paradigm for high-energy physics data analysis that simultaneously optimizes discriminative power and physical consistency. The architecture' s modular design permits direct extension to other quantum chromodynamics processes, providing critical technical infrastructure for next-generation intelligent particle physics analysis systems. Key implementation details and ablation studies are presented to elucidate the relative contributions of each architectural component.

Keywords: Z boson decay, Kolmogorov-Arnold Network, Residual Network, Deep Learning

Introduction

Z bosons, as mediators of the weak interaction in the Standard Model, hold irreplaceable value in particle physics research [1]. The lepton decay modes of the Z boson, $Z\to\mu^+\mu^-$ (Zmumu) and $Z\to e^+e^-$ (Zee), provide a critical window for exploring fundamental particle interaction mechanisms [2]. These processes are not only central to testing the precision of the Standard Model [3] but may also offer clues to new physical phenomena, such as new particles or interaction mechanisms beyond the Standard Model [4].

The challenge of distinguishing similar decay modes, such as Zmumu and Zee, remains a bottleneck in experimental analysis. As the data produced by large-scale facilities like the Large Hadron Collider (LHC) grows exponentially [5], traditional analysis methods face significant challenges in handling high-dimensional, nonlinear data and fully exploiting potential physical insights. Therefore, the development of an automated classification tool that combines high precision with physical interpretability has become an urgent need in particle physics research [6]. Such tools could not only enhance the efficiency of Standard Model verification but may also provide crucial data support for the discovery of new physical phenomena through precise identification of decay modes.

In this study, we use the Z boson dataset from the CERN Open Data repository, which is primarily derived from datasets contributed by McCauley and Thomas. These datasets are based on the Run2011A primary datasets, SingleElectron, SingleMu, DoubleElectron, and DoubleMu (https://opendata.cern.ch/record/545). Using this dataset, we propose a novel classification model, KAN-ResCNN, which integrates deep learning with prior physical knowledge [7]. This model innovatively combines the Kolmogorov-Arnold Network (KAN) [8], Residual Network (ResNet) [9], and Convolutional Neural Network (CNN) [10], aiming to address the efficiency and interpretability shortcomings of traditional methods in classifying Z boson decay modes [11].

Specifically, KAN-ResCNN employs a dual-branch architecture: the ResNet branch extracts deep physical features via residual structures, the CNN branch captures local feature correlations, and the KAN module applies a theory-driven nonlinear mapping for classification decisions. This architecture not only enhances the model' s ability to represent complex physical laws but also provides physical interpretability through the visualization capability of KAN [12]. Experimental results show that, on the CERN Open Dataset, KAN-ResCNN achieves an accuracy of 94.73% and an F1 score of 94.75% in the binary classification task of Zmumu versus Zee, significantly outperforming traditional models and other classification models for Z boson decay modes [13]. The contributions of this study can be summarized as follows: (1) The first particle physics data analysis model integrating KAN with a deep residual-convolutional architecture is proposed; (2) A dual-branch design is employed to balance physical feature extraction with data-driven learning; (3) A synergistic improvement in classification accuracy and interpretability is achieved. This work provides an efficient and reliable analytical tool for high-energy physics experiments and offers methodological support for the intelligent transformation of particle physics research. Future work will extend this model to other particle decay modes, such as Higgs boson two-photon decays, for identification tasks [14].

II. Related Work

A. The Evolution of Residual Networks in Particle Physics

Residual Networks (ResNet) have revolutionized deep learning through their intrinsic skip-connection mechanism, which effectively mitigates gradient degradation in deep architectures while enabling stable training of networks with more than 100 layers [15]. This paradigm has recently permeated high-energy physics (HEP) research, particularly for analyzing detector-generated one-dimensional feature vectors. Pioneering applications include Ref. [16], which implemented transfer learning through fine-tuning of ResNet-50 architectures on LHC-generated 2D histograms (signal+background vs. background-only configurations), demonstrating 90.2% classification accuracy for dark matter signatures at signal-to-background ratios as low as 0.008. Concurrently, Ref. [17] developed a dual-ResNet framework for plasma jet dynamics prediction, leveraging computational fluid dynamics (CFD) simulations of multi-arc plasma spraying systems. Their cascaded architecture achieved relative errors below 3.5% in particle velocity/temperature predictions, reducing computational costs by two orders of magnitude compared to conventional CFD solvers—a critical advancement for real-time plasma diagnostics.

While promising, current ResNet adaptations in High Energy Physics (HEP) reveal several limitations: they primarily rely on pretrained vision models instead of physics-optimized architectures, underutilize domain knowledge in network design, and lack explicit mechanisms for enforcing theoretical constraints. To address these issues, we have developed a modified ResNet architecture that employs 1D convolutional kernels specifically optimized for collider detector outputs [18]. Through a parallel design with a CNN branch, the model extracts both deep physical laws and local feature correlations [19]. Additionally, the introduction of the KAN module provides a theory-driven nonlinear mapping, compensating for the traditional ResNet' s limitations in embedding physical prior knowledge [20]. As shown in Fig. 1 [Figure 1: see original paper], the input data first passes through the input layer, where local features are extracted via a convolutional layer. Subsequently, the data undergoes normalization and nonlinear transformation through a batch normalization layer followed by a ReLU activation function [21]. The data then sequentially passes through four residual blocks (illustrated in Fig. 2 [Figure 2: see original paper]), with each block consisting of convolutional layers, batch normalization layers, ReLU activations, and skip connections, progressively extracting deeper-level features. Finally, a flattening layer transforms the multi-dimensional data into a one-dimensional vector, preparing it for subsequent fully connected layers or the output layer [22]. This architecture effectively addresses the vanishing gradient problem by improving gradient flow during backpropagation, which allows for the successful training of very deep models, especially in tasks like image classification and object recognition [23].

B. Local Feature Extraction of Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have demonstrated significant advantages in processing raw data in particle physics due to their local perception and parameter sharing characteristics [24]. By stacking convolutional and pooling layers to extract spatial features, CNNs have achieved remarkable success in tasks such as jet classification [25] and event reconstruction [26]. However, traditional CNNs still have two major limitations in particle physics analysis: first, the local perception property of CNNs causes the model to overly focus on spatial patterns (such as track topology) while neglecting global physical laws (such as statistical properties of the lepton pair invariant mass); second, the "black-box" nature of CNNs makes their decision-making process difficult to associate with Standard Model theory, limiting the interpretability of scientific discoveries [27].

To address these issues, this study introduces three improvements to the CNN architecture: first, replacing traditional 2D convolutions with 1D convolutions, which allows direct processing of one-dimensional time-series features (such as track curvature and energy deposition distributions) produced by detectors, thereby avoiding information loss caused by transformation to 2D images [28]; second, using adaptive average pooling to map input features of varying lengths to fixed-dimensional outputs, ensuring compatibility with subsequent modules [29]; and third, designing parallel CNN and ResNet branches, where the CNN branch focuses on local pattern extraction (such as track topology and cluster shape), while the ResNet branch captures deep features through residual connections. Ultimately, the features are complemented through tensor concatenation, effectively balancing local perception with global pattern modeling. The convolutional neural network architecture employed in this study is shown in Fig. 3 [Figure 3: see original paper]. The network begins with an input layer, which receives external data. This is followed by a convolutional layer with 32 filters, designed to extract local features from the input data. The ReLU activation function is applied to introduce nonlinearity into the model. Batch normalization is then performed to standardize the mean and standard deviation of each feature channel, enhancing model stability. A max-pooling layer reduces dimensionality while preserving the most significant features. Another convolutional layer follows, utilizing 64 filters to extract more complex and higher-level features, with ReLU activation and batch normalization applied to the output. Subsequently, an adaptive average pooling layer captures global features, providing a fixed-size output. Finally, a fully connected layer maps the extracted features to the final output space, facilitating the decision-making process.

C. Design of the Kolmogorov-Arnold Network Module

The Kolmogorov-Arnold Network (KAN) is a neural network architecture based on the Kolmogorov-Arnold representation theorem [8], designed to approximate high-dimensional functions through piecewise polynomial approximations while maintaining both model interpretability and computational efficiency [30]. Tra-

ditional neural networks, such as Multi-Layer Perceptrons (MLPs), although possessing powerful function approximation capabilities, suffer from the "black-box" nature [31], making their decision-making process difficult to correlate with physical laws, which limits their applicability in scientific discovery. KAN enhances interpretability significantly by incorporating learnable spline functions and basis functions, while preserving computational efficiency. The basic structure of the KAN network, as depicted in Fig. 4 [Figure 4: see original paper], comprises multiple hierarchical layers through which data flows. In each layer, the input data undergoes processing by a specific set of functions, which then transmit the processed output to the subsequent layer. The activation function used at each node is illustrated, with the B-spline function being employed in this network. The role of the B-spline function is critical, as its graphical representation demonstrates how the "grid extension technique" facilitates the transition between coarse and fine grids. This method, in conjunction with the dynamic adjustment of the network's multi-layered structure and activation functions, enables the network to effectively handle high-dimensional data. Such a design not only accommodates varying data resolutions but also optimizes overall performance by fine-tuning the precision of the activation functions.

1. Core Design of KAN The Kolmogorov-Arnold representation theorem forms the theoretical foundation of KAN. This theorem asserts that any multivariate continuous function $f(x_1, x_2, ..., x_n)$ can be represented as a sum of a finite number of univariate functions: $f(x_1, x_2, ..., x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \lambda_{q,p}(x_p) \right)$. In this context, Φ_q is the outer function responsible for combining the outputs of lower-dimensional functions into the final result, while $\lambda_{q,p}$ represents the inner functions that handle each input dimension individually. Based on this, KAN decomposes the high-dimensional function into a combination of multiple lower-dimensional functions, achieved through the following key components:

Piecewise Polynomial Approximation: KAN utilizes B-spline basis functions to perform piecewise polynomial approximation on the input data. B-spline basis functions have local support and smoothness, enabling them to effectively capture the local features of the input data. The mathematical expression for a B-spline basis function is: $B_{i,k}(x) = \frac{x-t_i}{t_{i+k}-t_i} B_{i,k-1}(x) + \frac{t_{i+k+1}-x}{t_{i+k+1}-t_{i+1}} B_{i+1,k-1}(x)$, where $B_{i,k}(x)$ is the i-th B-spline basis function of degree k, and t_i is the knot vector that defines the support interval of the spline function. This expression is recursively defined, with lower-degree spline functions (of degree $k-1$) being linearly combined to generate higher-degree spline functions (of degree k).

Fusion of Basis Functions and Spline Functions: KAN combines basis functions and spline functions to capture both the global trends and local details of the input data. The basis functions compute global features through linear transformations, while the spline functions calculate local features via B-spline basis functions and learnable weights. This design allows

KAN to model both global patterns and local structures simultaneously. The output can be expressed as: $y = \text{BaseOutput}(x) + \text{SplineOutput}(x)$, where $\text{BaseOutput}(x) = \text{SiLU}(W_{\text{base}} \cdot x)$ and $\text{SplineOutput}(x) = \sum_i w_i \cdot B_i(x)$. The BaseOutput branch of the basis function computes global features by applying a linear transformation (W_{base}) followed by the SiLU activation function, thereby capturing the overall trend of the input data. The SplineOutput branch of the spline function computes local features by utilizing B-spline basis functions $B_i(x)$ and learnable weights w_i, effectively capturing the detailed patterns within the input data.

Dynamic Grid Update: To enhance the model' s adaptability to the distribution of input data, KAN introduces a dynamic grid update mechanism. This mechanism adjusts the grid points of the spline function dynamically based on the distribution of the input data, ensuring the model can adaptively capture the key features of the data. The grid update formula is: $\text{grid} = \text{grid}_{\text{uniform}} \cdot \tau + \text{grid}_{\text{adaptive}} \cdot (1 - \tau)$, where $\text{grid}_{\text{uniform}}$ represents a uniform grid, which is evenly distributed based on the range of the input data, and $\text{grid}_{\text{adaptive}}$ is an adaptive grid, which dynamically adjusts according to the actual distribution of the input data. τ is the grid adjustment parameter that controls the weighting between the uniform and adaptive grids.

2. Innovations in KAN Traditional implementations require the expansion of all intermediate variables to compute different activation functions, leading to significant memory overhead. KAN reduces memory costs significantly by reformulating the computation process, transforming the activation function computation into a linear combination of basis functions. The original KAN implementation used L1 regularization on intermediate tensors for nonlinear operations, which is incompatible with the reformulated computation. In this revised approach, the L1 regularization is applied to the weights, and the formula is as follows: $L_{\text{reg}} = \lambda_1 \sum |W_{\text{spline}}| + \lambda_2 \sum p \log p$. The **Activation Term** $\sum |W_{\text{spline}}|$ represents the L1 regularization applied to the spline function weights, which serves to constrain the model' s complexity and prevent overfitting. The **Entropy Term** $\sum p \log p$ is the entropy regularization applied to the spline function weights, designed to enhance the sparsity and interpretability of the model. λ_1 and λ_2 are the regularization weights for the activation and entropy terms, respectively. The regularization loss function improves the generalization ability and interpretability of the model by constraining the distribution of the spline function weights.

III. Method

A. Overall Architecture Design

The proposed KAN-ResCNN model employs a dual-branch parallel architecture aimed at simultaneously capturing local patterns at the detector level and global physical laws. By integrating the deep feature extraction capability of Residual Networks (ResNet), the local pattern capture ability of Convolutional Neural

Networks (CNN), and the interpretable mapping of the Kolmogorov-Arnold Network (KAN), this model provides a comprehensive approach. As shown in Fig. 5 [Figure 5: see original paper], the input to the model is a one-dimensional feature vector representing the Z boson decay event, from which classification decisions are made through the following steps. The CNN branch extracts local physical features using 1D convolutional layers and adaptive average pooling, while the ResNet branch captures deep global features through the stacking of residual blocks. After the concatenation of features from both branches, they are fed into the KAN classifier, which leverages a combination of spline functions and basis functions to achieve a theory-driven nonlinear mapping.

B. Key Technological Innovations and Implementation Details

1. Dynamic Dimension Fusion and Adaptive Classification During the initialization phase, the model achieves flexible feature fusion by dynamically calculating the feature dimensions. The input data undergoes forward propagation, automatically computing the concatenated dimensions of the outputs from both branches. First, a random tensor representing the input is passed through the CNN branch to calculate the feature dimensions of that branch. Then, the ResNet branch processes the input data and computes the corresponding feature output. Ultimately, the total dimension of the feature fusion is obtained by summing the feature dimensions from both branches. This dynamic dimension computation mechanism enables the model to adapt to variations in input feature lengths, ensuring the robustness of feature concatenation. Finally, the fused features are processed by the KAN classifier, which performs theory-driven nonlinear mapping on the combined information. The mathematical formulation is as follows: $y = \mathrm{KAN}(\mathrm{Concat}(F_{\mathrm{CNN}}, F_{\mathrm{ResNet}}))$, where F_{CNN} and F_{ResNet} represent the feature vectors of the CNN and ResNet branches, respectively.

2. Residual Block Optimization Design In response to the characteristics of particle physics data, this study improves upon the traditional residual block structure. Firstly, the conventional two-dimensional convolutional residual block is replaced with a one-dimensional convolution, aiming to better align with the temporal signal characteristics of the detector. Secondly, to mitigate the common issue of vanishing gradients in deep networks, a shortcut connection is introduced within the residual link, thereby improving gradient stability. Additionally, a channel expansion strategy is employed, progressively increasing the number of channels at each layer, effectively enhancing the feature representation capacity. The forward propagation process of the residual block can be formally expressed as: $F(x) = \mathrm{ReLU}(f_2(f_1(x)) + S(x))$, where f_1 and f_2 represent convolutional layers, and $S(x)$ denotes the shortcut connection branch. This modified residual block structure significantly enhances the model' s ability to process particle physics data.

3. Physics-Inspired Feature Extraction In response to the characteristics of detector data, an optimized feature extraction module is designed in

this study, comprising both CNN and ResNet branches. In the CNN branch, two-stage 1D convolutions are employed to extract track topology features, capturing local patterns. Subsequently, global average pooling (AdaptiveAvgPool1d) is applied to compress the dimensionality of the features, generating fixed-dimensional feature representations. During model initialization, the output dimension of the CNN branch is computed using synthetic input data to accommodate varying feature lengths of the input. The ResNet branch, on the other hand, learns global physical patterns progressively by stacking four residual blocks. This design enables the model to fully leverage physics-inspired principles to uncover deeper patterns within the detector data.

IV. Experiments

A. Experimental Setup and Dataset Preprocessing

The experimental dataset employed in this study originates from the CMS experiment at the CERN-LHC, specifically derived from the "Datasets derived from the Run2011A SingleElectron, SingleMu, DoubleElectron, and DoubleMu primary datasets" curated by McCauley and Thomas. Our analysis focuses on the critical classification task of distinguishing between two key weak interaction decay channels of the Z boson: Zmumu and Zee. These channels exhibit pronounced physical similarities, presenting two primary classification challenges: (i) topological consistency in final-state particle signatures, and (ii) overlapping lepton-pair invariant mass distributions within the 60–120 GeV/c^2 range. The curated dataset comprises 20,000 high-confidence collision events, each systematically annotated with 10 fundamental physical observables: Run (the run number of the event), Event (the event number), pt1 and pt2 (the transverse momentum of the lepton, either a muon or an electron, in GeV), Q1 and Q2 (the charge of the lepton, either a muon or an electron), phi1 and phi2 (the phi angle of the lepton, either a muon or an electron, in radians), eta1 and eta2 (the pseudorapidity of the lepton, either a muon or an electron). These parameters collectively encode the kinematic characteristics of decay products and electromagnetic interaction properties. The intrinsic correlations among parameters, coupled with sub-detector-level measurement precision, establish a rigorous benchmark for evaluating machine learning classifiers in high-energy physics applications. Furthermore, the dataset undergoes standardized preprocessing to ensure cross-experimental consistency in physical parameterization. This meticulously processed dataset serves as a challenging testbed for developing advanced classification algorithms capable of addressing nuanced pattern recognition tasks in particle physics phenomenology.

The dataset was initially imported via the Pandas library in Python, followed by fundamental preprocessing operations. The categorical labels ("Zmumu" and "Zee") were numerically encoded (0 and 1) to support deep learning applications. The data preprocessing involved two key stages: dimensionality reduction via feature selection, along with outlier detection and normalization. Outlier analysis employed boxplot visualization (Fig. 6 [Figure 6: see original

paper]), where vertical axes display normalized feature distributions (horizontal range 0-1). Prolonged box lengths indicated dispersed distributions with potential outliers, defined mathematically as observations beyond Q1-1.5IQR or Q3+1.5IQR boundaries (IQR = interquartile range). Critical examination revealed significant outlier presence in pt1, pt2, Q1, and Q2 features. Empirical validation demonstrated substantial experimental impact from Q1/Q2 outliers, prompting their removal to enhance data reliability. For feature optimization, a correlation matrix visualization (Fig. 7 [Figure 7: see original paper]) was employed, where axes denote dataset features and chromatic intensity reflects Pearson correlation coefficients (-1 to 1). Features exhibiting absolute correlations exceeding 0.7 were identified as redundant, with strategic retention of single representatives from collinear groups. This approach simultaneously enabled identification of predictive features through their correlation patterns with key variables, while the matrix' s inherent symmetry validated dataset consistency. Post-feature elimination, MinMaxScaler standardization was implemented using the transformation: $X_{\text{norm}} = \frac{X-\mu}{\sigma}$, where μ and σ represent feature-specific mean and standard deviation, respectively, mapping values to [0,1] intervals. This preprocessing pipeline established a robust foundation for subsequent model development by ensuring dimensional consistency, eliminating noise, and maintaining feature interpretability while optimizing computational efficiency. The systematic approach effectively balanced dataset integrity with predictive utility, crucially enhancing model generalizability through reduced overfitting risks from redundant or anomalous data points.

B. Model Training and Evaluation

In the development of deep learning models, the processes of training and validation are essential for ensuring robust performance and generalization on unseen data. These steps are particularly vital in deep learning frameworks, as they help optimize the integration of various algorithms to achieve peak performance. For this study, we utilized the Z boson Dataset, which contains 20,000 events, each characterized by 10 features. The dataset was partitioned into three distinct subsets: 70% allocated for training, 10% for validation, and 20% for testing. This stratified division ensures that the model is trained and evaluated on non-overlapping data segments, providing a comprehensive assessment of its performance. Such an approach not only mitigates potential biases arising from imbalanced data distributions but also enhances the model' s ability to generalize across diverse data scenarios. By rigorously evaluating the model on separate validation and test sets, we can better understand its behavior and ensure its reliability in real-world applications.

1. Model Architecture and Training Configuration This work proposes a hybrid neural network architecture that integrates deep representation learning and interpretability. The core design is realized through a multimodal feature fusion mechanism, an interpretable classification head, and an optimization strategy. Firstly, the model adopts a parallel dual-branch structure for hetero-

geneous feature extraction. The local feature capturing branch is based on a 1D temporal convolutional neural network to capture the local kinematic correlations of particle decay products. Hierarchical module stacking is employed for feature learning, with global average pooling layers used to compress spatiotemporal features while retaining key physical patterns. The deep residual learning branch constructs a four-stage residual network, with each stage consisting of two basic ResNet modules. The network extracts multi-scale features through incremental channel growth and layered down sampling, alleviating the vanishing gradient problem. The dual-branch outputs are concatenated through channel-wise fusion and fed into an interpretable classification head, optimizing both local and global features. The interpretable classification head uses a Kolmogorov-Arnold representation theorem-based interpretable network (KAN), replacing traditional activation functions with learnable B-spline basis functions. The decision boundary is explicitly expressed through the parametric combination of basis functions. The network structure is automatically configured according to the test input, ensuring that the input dimension strictly matches the fused feature dimension. The output layer utilizes temperature-scaled Softmax to balance classification confidence and the need for physical interpretability.

In terms of optimization strategy, the training process adopts a quadruple regularization scheme to improve the model's generalization ability. First, a dynamic learning rate decay strategy based on an exponential scheduler is used to enable rapid convergence in the early stages of training, followed by fine-tuning in later stages. Second, the AdamW optimizer is employed with weight decay (weight decay=1$\times10^{-4}$) $to prevent over-parameterization of the weight matrix. The initial learning rate is set to 1\times10^{-3}$ and is adapted through the exponential scheduler. Implicit regularization is introduced through batch normalization layers in the CNN/ResNet branches and a 0.2 probability Dropout layer, which synergistically enhance feature robustness. During training, cross-entropy is used as the loss function, as described by the formula: $L = -\sum_i \sum_c y_{ic} \log(p_{ic})$, where p_{ic} represents the predicted probability that sample i belongs to class c. Furthermore, in terms of computational efficiency optimization, the batch size is set to 64 samples per batch, balancing GPU memory efficiency and gradient update stability. This architecture ensures high representational power of the deep model while meeting the interpretability requirements of particle physics experiments for model decision transparency, providing empirical evidence for hyperparameter optimization.

2. Performance Evaluation and Ablation Study In this study, to comprehensively evaluate the performance of the proposed ensemble learning model, several standardized evaluation metrics were employed, including accuracy, precision, recall, and F1 score (Table 1). Accuracy is one of the most commonly used metrics for assessing the performance of classification models, representing the proportion of correctly classified samples out of the total number of samples. The basic formula for accuracy is given by: Accuracy =

$\frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$. In this formula, TP (True Positive) refers to the number of samples correctly classified as the positive class; TN (True Negative) refers to the number of samples correctly classified as the negative class; FP (False Positive) refers to the number of samples incorrectly classified as the positive class; FN (False Negative) refers to the number of samples incorrectly classified as the negative class.

Precision is a metric used to assess the proportion of true positive samples among all samples predicted as positive by the model. The fundamental formula for precision is: Precision $= \frac{\text{TP}}{\text{TP+FP}}$, where TP (True Positive) refers to the number of samples correctly classified as the positive class and FP (False Positive) refers to the number of samples incorrectly classified as the positive class. Recall is a metric used to measure the proportion of actual positive samples correctly identified by the model. The basic formula for recall is: Recall $= \frac{\text{TP}}{\text{TP+FN}}$, where TP (True Positive) refers to the number of samples correctly classified as the positive class and FN (False Negative) refers to the number of samples incorrectly classified as the negative class. F1 score is the harmonic mean of precision and recall, combining both metrics to provide a more comprehensive evaluation of model performance, particularly in cases of class imbalance. The formula for F1 score is: F1 $= \frac{2\times(\text{Precision}\times\text{Recall})}{\text{Precision+Recall}}$. The value of the F1 score ranges from 0 to 1, with higher values indicating a better balance between precision and recall.

We selected several machine learning models (CNN, W-KNN, KD-KNN, KNN, XGB) from recent literature and online resources, with the aim of distinguishing between the two decay processes of the Z boson. A comparative analysis was conducted to verify the effectiveness and superiority of our method. As shown in TABLE 1, our model demonstrates a significant improvement in accuracy compared to the other models.

TABLE 1. The accuracy, precision, recall, and F1 score of each classifier.

Model	Precision	Recall	F1 score	Accuracy	TT (sec)
W-KNN	69.20%	69.07%	69.01%	69.05%	0.060
KD-KNN	68.18%	67.93%	67.80%	67.90%	0.053
KNN	66.93%	66.87%	66.82%	66.85%	0.055
XGB	79.30%	72.06%	86.07%	75.98%	0.460
CNN	81.87%	81.45%	81.36%	81.42%	1.334
KAN-ResCNN	96.90%	92.70%	94.75%	94.73%	2.678

To validate the effectiveness of the architectural design, we conducted an ablation study on the KAN-ResCNN model, with detailed results presented in TABLE 2 . In this study, we explored the impact of various combinations of the ResNet, CNN, and KAN modules on the model' s performance. Specifically, we first retained the ResNet branch while removing both the CNN and KAN components to evaluate the effectiveness of the deep residual network in feature

extraction and overall model performance. Next, we only kept the CNN branch, removing ResNet and KAN, to assess the CNN' s capability in local feature extraction. Subsequently, we combined ResNet with KAN and CNN with KAN to examine the role of the KAN module in enhancing model interpretability. By comparing the experimental results across different configurations, we found that the integration of KAN with both ResNet and CNN significantly improved the model' s discriminative ability, especially in terms of Precision and Recall, thereby substantially enhancing its overall performance. Consequently, the combination of ResNet, CNN, and KAN was identified as the optimal configuration, leading to significant improvements in both the accuracy and robustness of the model.

TABLE 2. An ablation study of each module in KAN-ResCNN for the classification of Zmumu and Zee decay modes, where denotes the addition of corresponding components or strategies.

ResNet	CNN	KAN	Precision	Recall	F1 score	Accuracy
			90.04%	79.34%	84.35%	84.90%
			79.30%	72.06%	86.07%	75.98%
			91.17%	84.50%	87.71%	87.85%
			88.78%	81.38%	84.92%	85.18%
			96.90%	92.70%	94.75%	94.73%

Fig. 8 [Figure 8: see original paper] presents three graphs that illustrate the model' s performance on different training metrics. The first plot, Loss Dynamics with Trendlines, has training epochs (ranging from 1 to 150) on the x-axis and loss values on the y-axis, which reflect the discrepancy between the model' s predictions and the true labels. Lower loss indicates better model performance. The blue solid line represents the training loss (Train Loss) across epochs, which initially starts high but decreases as the model learns, indicating the model' s adaptation to the training data. The green solid line shows the validation loss (Val Loss), which ideally should decrease in parallel with the training loss. However, after epoch 35, the validation loss begins to increase, signaling overfitting. The dashed trendlines are quadratic polynomial fits to the loss curves, illustrating the overall trend. The blue dashed line corresponds to the training loss trend, while the green dashed line represents the validation loss trend. The R^2 value, which measures the fit between the trendline and the actual data (ranging from 0 to 1), is 0.96 in this case, indicating that the trendlines account for 96% of the variation in the data.

The second plot, Accuracy Progression, also uses training epochs on the x-axis, with accuracy (ranging from 0 to 1, where higher values indicate better performance) on the y-axis. The blue solid line represents the training accuracy (Train Acc), showing an increase in accuracy as the model learns. The green solid line represents the validation accuracy (Val Acc), which reflects the model' s performance on unseen data. The red dots (Best Acc) mark the highest validation

accuracy achieved during training, indicating the model' s optimal performance and corresponding epoch. The trendline further highlights the long-term trend in accuracy progression. The third plot, F1-Score Evolution, shows training epochs on the x-axis and the F1 score (ranging from 0 to 1) on the y-axis. The blue solid line represents the training F1 score (Train F1), which indicates the model' s ability to balance the identification of both positive and negative samples. The solid green line shows the validation F1 score (Val F1), with the trendline revealing the overall direction of the evolution of the F1 score throughout the training process. These visualizations provide a comprehensive overview of the model' s performance across training and validation datasets, highlighting key metrics such as loss, accuracy, and F1 score, as well as trends indicating overfitting and optimal model performance.

V. Summary

This work proposes the KAN-ResCNN model, which addresses the challenge of classifying Z boson decay modes by achieving synergistic optimization of high accuracy and physical interpretability. Large-scale experiments based on the publicly available CERN dataset demonstrate that the model achieves 94.73% accuracy and 94.75% F1 score in the binary classification tasks of $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$, showing a significant improvement of 9.83% in accuracy over traditional machine learning models (e.g., XGBoost, SVM) and mainstream deep learning models (e.g., pure ResNet, CNN).

The core innovations of the model are as follows. First, the heterogeneous feature extraction architecture of the dual branch uses the residual structure of the ResNet branch to model deep physical laws, combined with the local convolution operations of the CNN branch to capture detector-level features. This design enables complementary enhancement of global physical laws and local feature representations. Ablation experiments show that the dual-branch design improves classification accuracy by 6.9% compared to a single-branch model. Second, the theory-driven KAN classification decision leverages the interpretability advantage of the Kolmogorov-Arnold representation theorem, transforming the black-box decision making of traditional neural networks into an explicit mapping based on spline functions and combinations of basis functions. The KAN module dynamically adjusts the weights of the features through a grid update mechanism, and its visualized decision path aligns closely with theoretical predictions of standard models, providing a new paradigm for the inversion of the physical mechanism. Finally, data-driven, particle-physics-oriented optimization introduces innovative methods such as dynamic dimension fusion, residual block channel expansion, and physics-inspired regularization. Experiments show that the improved residual block structure significantly improves the stability of the gradient, while the B-spline-based L1-entropy regularization strategy effectively mitigates overfitting, resulting in a noticeable reduction in generalization error on the test set.

The results of this study provide new tools for intelligent analysis in high-energy

physics experiments. The methodology is extendable to particle decay mode recognition tasks such as Higgs boson and top quark decay modes. Future work will focus on three aspects: extending the model to multi-decay channel classification, exploring dynamic modeling of cascade decay processes; incorporating attention mechanisms to strengthen the extraction of key physical features; and developing a KAN-based framework for physical law discovery, achieving a closed-loop analysis from data-driven to theory-driven approaches. This research offers technical support for the deep integration of particle physics and artificial intelligence, with significant scientific value and application prospects.

References

[1] P. Langacker, M.X. Additional Z bosons. *Phys. Rev. D* 45, 278-292 (1992). https://link.aps.org/doi/10.1103/PhysRevD.45.278

[2] V.A. Novikov, L.B. Okun, N.A. Rozanov, et al., Theory of Z boson decays. *Rep. Prog. Phys* 62, 1275 (1999). https://doi.org/10.1088/0034-4885/62/9/201

[3] K, T, Experimental verification of the standard model of particle physics. *Proc. Japan Acad., Ser. B* 97, 211-235 (2021). https://doi.org/10.2183/pjab.97.013

[4] S. Antusch, J.P. Baumann, E. Fernández-Martínez, Non-standard neutrino interactions with matter from physics beyond the Standard Model. *Nucl. Phys. B* 810, 369-388 (2009). https://doi.org/10.1016/j.nuclphysb.2008.11.018

[5] P. Collier, The technical challenges of the Large Hadron Collider. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci* 373, 20140044 (2015). https://doi.org/10.2307/24506083

[6] V. Olshevsky, Y. Olshevsky, A. Lalti, et al., Automated classification of plasma regions using 3D particle energy distributions. *J. Geophys. Res. Space Physics* 126, 2021JA029620 (2021). https://doi.org/10.1029/2021JA029620

[7] E. De Bézenac, A. Pajot, P. Gallinari, Deep learning for physical processes: incorporating prior scientific knowledge. *J. Stat. Mech. Theor. Exp* 2019, 124009 (2019). https://dx.doi.org/10.1088/1742-5468/ab3195

[8] Z.M. Liu, Y.X. Wang, S. Vaidya, et al., Kan: Kolmogorov-arnold networks. arXiv preprint arXiv:2404.19756 (2024). https://doi.org/10.48550/arXiv.2404.19756

[9] M. Kubu, P. Bour, CNN with residual learning extensions in neutrino high energy physics. *J. Phys. Conf. Ser* 1730, 012133 (2021). https://doi.org/10.1088/1742-6596/1730/1/012133

[10] S.Y. Chen, T. Wei, C, Zhang, et al., Quantum convolutional neural networks for energy analysis in high energy physics. *Phys. Rev. Res* (2022). https://doi.org/10.1103/PhysRevResearch.4.013231

[11] P.J. Sadowski, W. Daniel, P. Baldi, Searching for Higgs Boson Decay Modes with Deep Learning. In: Z. Ghahramani, M. Welling, C. Cortes, et al., (eds)

Advances in Neural Information Processing Systems, vol 27. (Curran Associates, Inc., 2014). https://proceedings.neurips.cc/paper-files/paper/2014

[12] N. Ranasinghe, Y. Xia, S. Seneviratne, et al., Ginn-kan: Interpretability pipelining with applications in physics informed neural networks. arXiv preprint arXiv:2408.14780 (2024). https://doi.org/10.48550/arXiv.2408.14780

[13] C.D. Sag, O. Sahin, Predicting Z Boson Decay Modes: Evaluating the Performance of Machine Learning and Deep Learning Techniques in Particle Physics. IEEE (UBMK) 17-22 (2023). https://doi.org/10.1109/UBMK59864.2023.10286786

[14] S.D. Lane, H.S. Lee, I.M. Lewis, Multi-photon decays of the Higgs boson at the LHC. arXiv preprint arXiv:2305.00013 (2023). https://doi.org/10.48550/arXiv.2305.00013

[15] D.X. Wu, Y.S. Wang, S.T. Xia, et al., Skip connections matter: On the transferability of adversarial examples generated with resnets. arXiv preprint arXiv:2002.05990 (2020). https://doi.org/10.48550/arXiv.2002.05990

[16] A. Celik, Exploring hidden signal: Fine-tuning ResNet-50 for dark matter detection. *Comput. Phys. Commun.* 305, 109348 (2024). https://doi.org/10.1016/j.cpc.2024.109348

[17] K. Bobzin, H. Heinemann, SR Dokhanchi, et al., Replication of Particle Trajectories in the Plasma Jet with Two Consecutive Residual Neural Networks. *J. Therm. Spray Technol.* 32, 1447-1464 (2023). https://doi.org/10.1007/s11666-023-01533-1

[18] W. Ling, K.J Pan, Z.Y Ren, et al., One-dimensional magnetotelluric parallel inversion using a ResNet1D-8 residual neural network. *Comput. Geosci.* 180, 105454 (2023) https://doi.org/10.1016/j.cageo.2023.105454

[19] M. Jogin, M.S. Madhulika, G.D. Divya, et al., Feature extraction using convolution neural networks (CNN) and deep learning. IEEE (RTEICT) 2319-2323 (2018). https://doi.org/10.1109/RTEICT42901.2018.9012507

[20] R.C. Yu, S. Wu, J. Gui, Residual kolmogorov-arnold network for enhanced deep learning. arXiv preprint arXiv:2410.05500 (2024). https://doi.org/10.48550/arXiv.2410.05500

[21] H. Ide, T. Kurita, Improvement of learning for CNN with ReLU activation by sparse regularization. In: 2017 International Joint Conference on Neural Networks (IJCNN), pp. 2684–2691 (2017). https://doi.org/10.1109/IJCNN.2017.7966185

[22] H. Gholamalinezhad, H. Khosravi, Pooling methods in deep neural networks, a review. arXiv preprint arXiv:2009.07485 (2020). https://doi.org/10.48550/arXiv.2009.07485

[23] R.U. Khan, X. Zhang, R. Kumar, et al., Evaluating the performance of resnet model based on image recognition. *Proc. In: Proceedings of the 2018 International Conference on Computing and Artificial Intelligence*, pp. 86–90 (2018). https://doi.org/10.1145/3194452.3194461

[24] Z.T. Huang, L.Y. Wang, Q.Y. Ge, et al., An intelligent fault diagnosis method for CNN-SVM circuit breaker based on quantum particle swarm optimization. *J. Phys. Conf Ser.* 2113, 012047 (2021). https://doi.org/10.1088/1742-6596/2113/1/012047

[25] S.H. Lim, M. Nojiri, Morphology classification. *Phys. Rev. D* 105, 014004 (2022). https://doi.org/10.1103/PhysRevD.105.014004

[26] J.Y. Araz, M. Spannowsky, Combine and conquer: ensemble event reconstruction with Bayesian neural networks. *J. High Energy Phys.* 2021, 1-23 (2021). https://doi.org/10.1007/JHEP04(2021)296

[27] S. Thais, P. Calafiura, G. Grigorios, et al., Graph Neural Networks in Particle Physics: Implementations, Innovations, and Challenges. arXiv preprint arXiv:2203.12852 (2022). https://doi.org/10.48550/arXiv.2203.12852

[28] C. Shimmin, Particle physics. arXiv preprint arXiv:2107.02908 (2021). https://doi.org/10.48550/arXiv.2107.02908

[29] A. Zafar, M. Aamir, N. Mohd Nawi, et al., A comparison of pooling methods for convolutional neural networks. *Appl. Sci.* 12, 8643 (2022). https://doi.org/10.3390/app12178643

[30] M. Révayová, C. Török, Piecewise approximation and neural networks. *Kybernetika* (2007). http://www.kybernetika.cz/content/2007/4/547

[31] J.E. Dayhoff, J.M. DeLeo, Artificial neural networks: opening the black box. *Cancer* S8, 1615-1635 (2001). https://doi.org/10.1002/1097-0142

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