

Exploring the power of KANs: Overcoming MLP limitations in complex data analysis

Jiwei Liu

Qingdao Institute of Software, College of Computer Science and Technology, China
University of Petroleum (East China), Shandong, Qingdao, 266580, China

2207020506@s.upc.edu.cn

Abstract. In the field of artificial intelligence, the requirements of models that can adeptly explore informative features of complex data sets has led to the emergence of advanced neural network architectures. Beyond the perceptron-based architectures that are currently in widespread use, a variety of innovative and cutting-edge designs have been proposed. This paper thoroughly explores the critical role of Kolmogorov-Arnold Networks (KAN) in overcoming the limitations of traditional Multilayer Perceptron (MLP) models, particularly highlighting KAN's significant advantages in handling complex nonlinear and high-dimensional data. By analyzing the mathematical foundations and neural network architecture of KAN, and comparing it with MLP, the paper demonstrates KAN's outstanding performance in time series analysis and image classification. The research indicates that KAN has distinct advantages in addressing high-dimensional nonlinear data. The paper also summarizes the current research progress on KAN and discusses its enormous potential in future machine learning and real-world applications, pointing out possible future research directions.

Keywords: Kolmogorov-Arnold network, Time series analysis, Multilayer perceptron, Image classification.

1. Introduction

An artificial neural network (ANN) is a computational model inspired by the biological neural networks found in the brain. It consists of layers of artificial neurons, where each neuron simulates a biological neuron by taking inputs, applying weights, and passing them through an activation function to produce an output. Through a process called learning, ANNs adjust these weights to perform tasks such as classification, pattern recognition, and prediction. The traditional ANN has limited representational power, only capable of solving linearly separable problems and unable to handle complex nonlinear relationships. Additionally, due to its simple structure, typically comprising only input and output layers without depth or hierarchy, it performs poorly on complex tasks. Researchers found that by introducing hidden layers and using nonlinear activation functions based on the perceptron, ANNs can solve problems that are not linearly separable. This led to the emergence of the Multilayer Perceptron (MLP), as they can model more complex patterns and features, thus increasing the depth and complexity of the model. Moreover, MLPs can handle complex nonlinear problems, which are common in classification and regression tasks.

In recent years, ANNs have gained widespread popularity and have proven to be invaluable for tasks such as classification, clustering, pattern recognition, and prediction across many disciplines. As a key component of machine learning (ML), ANNs have driven significant advances in areas such as speech recognition, natural language processing, and autonomous systems.

MLP is a typical representative of neural networks, the neurons in each layer of MLP are fully connected with all the neurons in the previous layer to form a fully connected structure. Additionally, MLP uses nonlinear activation functions in the neurons of the hidden layer and the output layer. And it can be trained by Backpropagation algorithm. Theoretically, MLP has the ability to approximate arbitrary continuous functions as long as the network has enough layers and neurons. This makes MLP widely used in the fields of classification, regression, pattern recognition, time series prediction and so on.

However, is the MLP the best structure in the field of neural networks? Does the MLP have no drawbacks? As MLPs continued to be applied and the demands of the real world increased, some of the limitations of MLPs became apparent.

Due to the fully connected structure of MLPs, where each neuron in a layer is connected to all neurons in the previous layer, a large number of parameters are required for training. When dealing with high-dimensional data, the number of parameters increases significantly. Because MLPs have numerous parameters, especially in high-dimensional data scenarios, they are prone to overfitting, which increases computational complexity. To avoid overfitting, regularization techniques, Dropout, and other methods are typically employed, but these approaches cannot completely solve the problem. Although MLP can theoretically approximate arbitrary continuous functions, in practice, it requires a very deep network and a large number of neurons, which leads to the complexity of training and the consumption of computational resources. In addition, the design of MLP relies more on experience and experimentation and lacks mathematical basis in explaining its ability to handle complex data.

To better address the various limitations of MLPs and more efficiently meet the increasing new demands in practical applications, a new neural network structure becomes especially important.

This paper proposes a very efficient alternative to MLP, Kolmogorov-Arnold Networks (KANs) [1]. Whereas MLPs are inspired by the universal approximation theorem, KANs are inspired by the Kolmogorov-Arnold representation theorem [2]. Like MLPs, KANs have fully-connected structures. However, for accuracy, much smaller KANs can achieve comparable or better accuracy than much larger MLPs in data fitting and PDE solving [3]. For application, KANs are shown to be useful collaborators helping scientists (re)discover mathematical and physical laws [3]. This paper will give an outlook on the theoretical architecture of KAN, specific applications, recent developments, and future directions of KAN.

Through this paper, readers will have a comprehensive and deep understanding of the network structure of KAN, the current status of applications, cutting-edge advances, existing major research results, and future development trends. Readers can easily identify the unsolved problems and research gaps in the current research on KAN, providing new directions for subsequent research and applications.

2. Theoretical Foundations

2.1. Kolmogorov-Arnold Theory

Kolmogorov pointed out that any continuous multivariable function can be represented as a superposition of a finite number of continuous univariate functions [4].

$$f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q\left(\sum_{p=1}^n \phi_{q,p}(x_p)\right) \quad (1)$$

Here, $\phi_{q,p}$ are univariate functions mapping each input variable x_p such $\phi_{q,p}: [0, 1] \rightarrow \mathbb{R}$, and $\Phi_q: \mathbb{R} \rightarrow \mathbb{R}$, univariate functions. This theorem's discovery has significant implications for neural networks and mathematics.

It demonstrates that, even in high dimensional spaces, smaller univariate functions can be used to generate complicated multivariate functions, which simplifies the function representation problem. Additionally, it offers a novel way to approximate multivariate functions by the superposition of univariate functions. This strategy broadens the applicability of approximation theory and offers fresh concepts and techniques for handling high-dimensional approximation problems.

More flexible neural network design is theoretically supported by the Kolmogorov-Arnold theorem. The flexibility and expressiveness of the model are increased by KAN, which uses learnable univariate functions in place of fixed linear parameters, which are the weights in typical neural networks. It makes neural networks more adept at handling intricate nonlinear interactions and proficient at tasks like symbolic regression, partial differential equation solution, and data fitting.

2.2. Structural Comparison

While MLPs have fixed activation functions on nodes ("neurons"), KANs have learnable activation functions on edges ("weights"). KANs have no linear weights at all -- every weight parameter is replaced by a univariate function parametrized as a spline [5].

In traditional MLPs, neurons use fixed activation functions, which remain unchanged during the training process. This limits the model's ability to capture complex and diverse data patterns, especially when dealing with multimodal data. Fixed activation functions struggle to describe the complex relationships between different modes, leading to a decline in model performance.

By parameterizing each weight as a spline function, KANs can dynamically adjust and flexibly perform nonlinear transformations based on input data. This flexibility allows KANs to better identify subtle differences in the data and adopt different nonlinear strategies in different data regions. The learnable activation functions enable KANs to adaptively adjust weights during training, enhancing the model's generalization ability and its capacity to handle complex and diverse data, thereby improving prediction accuracy.

3. KAN-based Applications

After reading about KANs, their operation, and possible practical uses, one can see how widely applicable KANs are to a variety of fields, such as real-time data processing, engineering, and scientific research. In fact, problems like high-dimensional data analysis, symbolic regression, and complex function approximation are especially well-suited for KANs. These tasks are relevant in domains like financial forecasting, where KANs can accurately predict market movements and describe nonlinear relationships. KANs are also excellent at solving partial differential equations (PDEs), which makes them a useful tool for simulating and optimizing complex systems in engineering and science [6].

Moreover, KANs have demonstrated potential in continuous learning, enabling models to learn new information while retaining previously learned knowledge. This feature is essential for applications that need to be updated and improved all the time, like software with adaptive user interfaces, real-time infrastructure monitoring, and automated trading systems.

In conclusion, KANs offer a flexible and effective method for addressing a broad variety of challenging issues in diverse fields, and they have the potential to completely transform a number of sectors by improving the precision, interpretability, and flexibility of predictive models.

3.1. KAN-based Time Series Analysis

Time series analysis is essential for forecasting future observations by analyzing historical data, such as predicting climate changes from past meteorological trends. It also helps in detecting anomalies, vital for cybersecurity, equipment failure detection, and medical monitoring, where early detection of irregularities in patient data can prevent health risks. Additionally, time series analysis explores causal relationships between variables, crucial for hypothesis testing and policy evaluation, and optimizes resource utilization in fields like traffic management by analyzing historical data for better future planning.

In a previous work, researchers find that KANs can leverage their adaptive activation functions for enhanced predictive modeling [7]. Following that researchers have performed an analysis of KANs for satellite traffic forecasting. The results highlighted several benefits of KANs, including superior forecasting performance and greater parameter efficiency [7]. In addition, researchers find that in meteorological data, KANs can accurately capture seasonal and trend changes, dynamically adapting to changes in data patterns during time series analysis. In the field of energy consumption forecasting, KANs are used to predict demand for electricity or other forms of energy, optimize scheduling, and reduce energy costs. In financial markets, KANs can capture complex nonlinear fluctuations in trading volumes and stock prices, providing more accurate predictions for investment strategies.

Previous models for time series analysis have several shortcomings: (1) Linear Assumptions: Traditional time series models, such as Autoregressive Integrated Moving Average Model (ARIMA), are based on linear assumptions, presuming future values are a linear combination of past values. However, real-world data often exhibit complex nonlinear relationships, such as those seen in financial markets and meteorological data, challenging the accuracy of these models. (2) Fixed Structure: Traditional models typically have a fixed structure that cannot dynamically adjust to changes in data, leading to a decline in forecasting performance when data patterns shift. (3) Challenges with Multivariate Data: Traditional models struggle with multivariate data, as increasing the number of variables raises model complexity and computational cost, making it difficult to identify complex inter-variable relationships. (4) Dependency Handling Limitations: Traditional models are limited in handling long-term and short-term dependencies, usually managing only short-term dependencies effectively. Increasing model complexity to address long-term dependencies can lead to overfitting.

To address these issues, researchers have proposed applying KANs in the field of time series forecasting, achieving remarkable results. KANs have several advantages: (1) Handling Nonlinear Relationships: KANs are capable of representing complex multivariate continuous functions as compositions of univariate functions, making them particularly effective in capturing nonlinear relationships in time series data, especially in multivariate forecasting like meteorological variables or stock prices. (2) Adaptive Optimization: KANs can automatically adjust network structure and activation functions based on input data patterns, allowing for adaptive optimization across different phases, preventing overfitting or underfitting, and improving robustness and accuracy in predictions.

Interpretability: KANs offer greater interpretability in time series forecasting by analyzing parameter changes in univariate functions, providing an intuitive understanding of the model's decision-making process, crucial in fields like financial forecasting and medical monitoring. (3) Robustness to Noise: KANs demonstrate strong robustness when dealing with noise and outliers, maintaining high predictive accuracy and stability even in the face of sudden events or unstable fluctuations, making them reliable in rapidly changing markets or complex environments.

Although KANs have demonstrated many advantages in the field of time series analysis, they still present some drawbacks and challenges: (1) Training Instability: KANs can experience instability during the training process, such as gradient explosion or vanishing, especially when handling time series data with long-term dependencies. This can make the model difficult to converge, ultimately affecting prediction accuracy. (2) Resource-Intensive Training: KANs require continuous optimization of their learnable univariate functions during training, resulting in relatively long training times and significant computational resource consumption, making them less favorable for large-scale time series data. (3) Dependency on Parameter Tuning: The performance of KANs is highly dependent on parameter settings. Selecting the appropriate parameters, such as activation functions, network layers, and nodes, often requires extensive experimentation and fine-tuning, increasing the model's complexity and usage difficulty.

To address these drawbacks, this paper discusses potential improvements and applications of KANs in the field of time series analysis. Firstly, optimizing algorithms and introducing regularization techniques will enhance the model's training stability, particularly when dealing with long-term dependencies. Secondly, developing more efficient computational methods or hardware accelerators will help reduce training time and computational resource consumption, making KANs more suitable

for large-scale datasets. Lastly, the introduction of automated hyperparameter optimization techniques will simplify the parameter selection process, reduce the complexity of use, and make KANs more accessible for widespread application.

3.2. KAN-based Image Classification

Image classification plays a crucial role in scientific research, particularly in the fields of computer vision and artificial intelligence. By assigning images to specific categories and aiding in the automatic recognition and analysis of image content, it has found extensive applications in various domains such as medical image analysis, remote sensing, autonomous driving, and security surveillance. By enhancing classification accuracy, image classification technology has facilitated scientific discoveries and technological advancements, providing essential tools for solving complex visual tasks. Moreover, image classification has driven the development of deep learning technologies and has contributed to the progress of other artificial intelligence applications.

Additionally, in some studies and experiments, KANs have been effectively used in the hyperspectral image classification, which is essential for military, agricultural, and environmental remote sensing applications [8]. KANs, by analyzing hyperspectral images, can help monitor the spread of pollutants, vegetation health, and changes in water quality, providing scientific evidence for environmental protection. In agriculture, KANs are used for monitoring crop types and pest infestations, helping farmers make better planting decisions and increase yields. These applications demonstrate the advantages of KANs in handling high-dimensional and complex spectral information and support precise analysis in various practical fields.

Previous models for image classification have several shortcomings. (1) Previous models have insufficient capacities for feature extraction. The problem arises when large numbers of photos are concerned. It becomes a too difficult problem to find features from them [9]. (2) The adaptability of models is insufficient. Convolutional Neural Networks (CNNs) and vision transformers (ViTs) have shown excellent capability in complex hyperspectral image (HSI) classification. However, these models require a significant number of training data and are computational resources [10]. (3) Conventional methods have inadequate handling of nonlinear relationships. Traditional methods mainly relied on handcrafted features and linear models, which struggled to capture the complex nonlinear relationships within image data, limiting classification accuracy.

To address these existing issues, KANs have several advantages. (1) Fast Convergence. KAN and MLP achieved the same accuracy in the first epoch. In the second epoch, KAN's accuracy quickly reached 96% and remained stable until the end. However, MLP's accuracy only reached 95% in the second epoch and increased to 96% in the third epoch. Specific experimental images can be found [9]. This revealed that KAN attained high accuracy slightly more quickly than MLP, but both ultimately reached similar accuracy, suggesting there is not a significant difference between the two for remote sensing classification tasks [9]. (2) Efficient parameter utilization. KAN can achieve high classification accuracy while substantially reducing the number of training parameters required. This efficiency makes the model less heavy, which makes it better for situations where computational resources are limited. (3) Reduced overfitting risk. Due to KAN's parameter efficiency and smaller architecture, it is less prone to overfitting compared to traditional MLPs, especially when handling high-dimensional data. (4) Improved interpretability and precision. KAN makes use of learnable activation functions that are based on wavelets or B-splines, which are more effective at capturing intricate spectral-spatial patterns. Improved interpretability and increased classification accuracy are the outcomes of this, particularly in hyperspectral image classification tasks. (5) Robust generalization capability. KAN can easily conform to intricate data structures. Its adaptable activation functions provide to better results across a variety of datasets, especially when managing intricate spectral-spatial correlations, by efficiently capturing multi-scale information in the data.

Although KANs have demonstrated many advantages in the field of image classification, they still present some drawbacks and challenges: (1) Interpretability limitations. Despite these results, this study acknowledges several shortcomings. One major limitation is the lack of evidence about the

interpretability of KAN layers, which is crucial for understanding the decision-making process of the model [9]. (2) Model complexity. KAN has a somewhat complicated structure, especially when learning activation functions such as B-splines or wavelets [11]. This makes developing and putting the model into practice more challenging. Furthermore, as comparison to simpler models, this complexity could make KAN less intuitive to use and more difficult to implement in specific application circumstances.

This experiment indicates that KAN layers can achieve high accuracy with fewer training epochs, even when the number of nodes is significantly reduced and this proved the efficiency of the KAN layer [10]. Future research should address these issues by improving KAN's interpretability and investigating more effective integration solutions for various remote sensing applications [9].

4. Discussion

4.1. Current Research Progress

Model Extensions: While depth-2 representations were the main focus of early KAN research, more recent work has expanded the network's structure to accommodate arbitrary widths and depths. This enhanced version, also known as B-Spline KAN, increases the expressiveness and versatility of the model by using B-Spline functions as learnable activation functions.

Application Domains: KAN has been effectively used in a number of domains, including as financial market analysis, hyperspectral image classification, and time series forecasting. It has demonstrated notable benefits, especially when managing multivariate dependencies and intricate nonlinear interactions. For instance, KAN efficiently collects both spectral and spatial data in hyperspectral image classification, leading to greater classification accuracy.

Performance and Challenges: KAN has trouble during training even if it performs well in a variety of applications. For instance, when working with long-term dependencies, the intricate structure of KAN may result in problems like gradient ballooning or disappearing. KAN can only be used on large-scale datasets because it takes a lot of processing power and lengthy training periods. Researchers are looking into automated hyperparameter tweaking, regularization strategies, and optimization algorithms to overcome these issues.

Comparing KAN with Other Models: Research indicates that in some applications, particularly those where high accuracy and interpretability are essential, KAN performs better than conventional MLPs and CNNs. But parameter selection plays a major role in how well KAN performs, especially when it comes to selecting activation functions and network design, which makes using the model more challenging.

4.2. Future Prospect

Creation of Hybrid Models: While KAN has demonstrated impressive performance in several applications, hybrid models might be created by merging it with other deep learning models (like recurrent or convolutional neural networks). By combining the advantages of several network topologies, these hybrid models may be able to enhance performance on intricate datasets. Future studies could examine the efficient integration of KAN with other models and assess their performance on a range of tasks.

Extension of Application Domains: Although KAN has shown promise in domains such as image classification and time series forecasting, it remains possible to extend its application to other fields. Subsequent investigations may examine the application of KAN in developing domains including autonomous driving, genetic data analysis, and natural language processing, evaluating its efficacy in these intricate situations. KAN might have even more usefulness and influence if its application domains were expanded.

Integration of Wavelets and additional Functions: Wavelet-based KAN variations have been studied so far; future study could look into ways to integrate additional functions (such polynomial or Fourier transformations) to increase the capabilities of KAN. KAN might demonstrate increased versatility and

adaptability when managing a broader range of jobs and data types by integrating several mathematical techniques.

5. Conclusion

This paper provides a comprehensive and in-depth exploration of KAN, highlighting its advantages in handling complex nonlinear data through a comparison with traditional MLP. The paper elaborates on the mathematical foundations of KAN, its neural network architecture, and its application in fields such as time series analysis and image classification. As an emerging neural network architecture, KAN excels in processing high-dimensional, complex nonlinear, and dynamic data through learnable activation functions and parameterized spline functions. It effectively addresses nonlinear relationships in financial market fluctuations and meteorological data and is capable of extracting and distinguishing high-dimensional features in hyperspectral image classification. Based on the Kolmogorov-Arnold representation theorem, KAN can decompose complex multivariable functions into combinations of univariate functions, thereby better capturing nonlinear relationships in data. This makes KAN suitable for complex economic models, nonlinear dynamic behavior in physical systems, and multivariable analysis in climate change predictions or financial markets.

However, given that KAN is relatively new, it still faces certain limitations, such as insufficient interpretability and high computational demands. Future research should focus on improving KAN's interpretability, expanding its application domains, and exploring hybrid models. The paper suggests that as these challenges are gradually addressed, KAN is likely to gain broader and deeper applications across various fields, further advancing the development of machine learning.

References

- [1] Kolmogorov, A. N. (1957). On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. In *Doklady Akademii Nauk*, 114(5), 953-956.
- [2] Braun, J., & Griebel, M. (2009). On a constructive proof of Kolmogorov's superposition theorem. *Constructive approximation*, 30, 653-675.
- [3] Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., et al. (2024). Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756*.
- [4] Kolmogorov, A. N. (1961). On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables. *American Mathematical Society*.
- [5] Cheon, M. (2024). Demonstrating the efficacy of kolmogorov-arnold networks in vision tasks. *arXiv preprint arXiv:2406.14916*.
- [6] Wang, Y., Sun, J., Bai, J., Anitescu, C., Eshaghi, M. S., et al. (2024). Kolmogorov Arnold Informed neural network: A physics-informed deep learning framework for solving PDEs based on Kolmogorov Arnold Networks. *arXiv preprint arXiv:2406.11045*.
- [7] Vaca-Rubio, C. J., Blanco, L., Pereira, R., & Caus, M. (2024). Kolmogorov-arnold networks (kans) for time series analysis. *arXiv preprint arXiv:2405.08790*.
- [8] Seydi, S. T. (2024). Unveiling the Power of Wavelets: A Wavelet-based Kolmogorov-Arnold Network for Hyperspectral Image Classification. *arXiv preprint arXiv:2406.07869*.
- [9] Singh, A., & Singh, P. (2020). Image classification: a survey. *Journal of Informatics Electrical and Electronics Engineering*, 1(2), 1-9.
- [10] Jamali, A., Roy, S. K., Hong, D., Lu, B., & Ghamisi, P. (2024). How to Learn More? Exploring Kolmogorov-Arnold Networks for Hyperspectral Image Classification. *arXiv preprint arXiv:2406.15719*.
- [11] Seydi, S. T. (2024). Unveiling the Power of Wavelets: A Wavelet-based Kolmogorov-Arnold Network for Hyperspectral Image Classification. *arXiv preprint arXiv:2406.07869*.