# A STUDY ON THE APPLICATION OF KOLMOGOROV–ARNOLD NETWORKS FOR RENEWABLE ENERGY FORECASTING

**JUAN YAN [1], YUHAO GAO [1], WENPING CAO [1*] CUNGANG HU [1]**

[1] School of Electrical Engineering and Automation Anhui University Hefei, China
E-MAIL: juanyan@ahu.edu.cn, z24301213@stu.ahu.edu.cn, wpcao@ahu.edu.en, hcg@ahu.edu.cn

**Abstract:**

**With the widespread adoption of clean energy technologies, accurate forecasting of renewable energy outputs has become increasingly critical. This study explores the application of Kolmogorov–Arnold Networks (KAN) in the domain of renewable energy prediction. KAN, inspired by the Kolmogorov–Arnold representation theorem, replace traditional linear weights in multilayer perceptrons (MLPs) with spline functions, enhancing model interpretability. Unlike conventional black-box models, KAN offer significant explanatory capabilities. We conducted experiments using real-world wind and photovoltaic power output data from Shandong Province, China. The results demonstrate that KAN outperform benchmark models, including MLPs and Long Short-Term Memory (LSTM) networks, in most scenarios, while also providing superior interpretability. These findings suggest that KAN hold substantial potential for practical applications in renewable energy forecasting.**

**Keywords:**

**Neural Networks; Intelligent Network; Applications; Kolmogorov–Arnold Networks (KAN); Renewable Energy Forecasting; Wind Power Prediction; Photovoltaic Power Forecasting**

## 1. Introduction

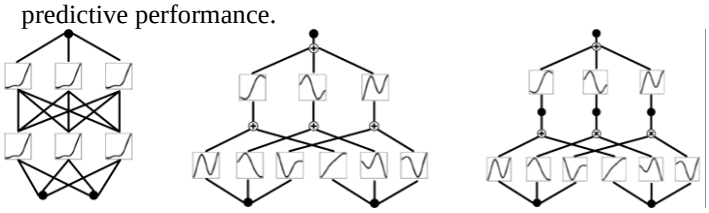
With the rapid development of the economy and the continuous improvement of living standards, the demand for energy has been steadily increasing. However, the use of traditional fossil fuels, such as coal, oil, and natural gas, releases pollutants that damage the environment and contribute to global warming [1]. Due to their non-renewable nature and limited reserves, the excessive exploitation of fossil fuels will eventually lead to the depletion of energy resources. Therefore, addressing the energy crisis and environmental issues necessitates the development and utilization of renewable energy sources [2]. Wind and solar energy are clean, widely distributed forms of renewable energy that have attracted global attention and are considered competitive and viable alternatives to fossil fuels [3][4].However, the intermittency and randomness of these renewable energy sources significantly impact the security and stability of power grids. Additionally, challenges in dispatching, managing, and optimizing renewable energy integration at various grid scales are major barriers to achieving high penetration levels [5]. Therefore, improving the prediction accuracy of renewable energy generation is a key solution to these issues.

In wind power forecasting, several types of modeling approaches are generally employed: physics-based methods, statistical techniques, artificial intelligence–driven models, and hybrid models that combine these approaches. Guo N.Z. et al. proposed a short-term wind power prediction neural network that accounts for wake effects, in which certain network nodes are determined by an analytical wake model to enhance the model's short-term accuracy [6]. Kim Y. and Hur J. introduced a statistical ensemble forecasting model composed of three combined prediction methods and applied it to a wind farm in Jeju Island, South Korea [7]. Kisvari A. et al. developed a novel deep learning network based on gated recurrent units (GRU) and demonstrated its superior accuracy and robustness [8]. Ma Z. and Mei G. proposed a hybrid attention-based deep learning approach that combines convolutional neural networks (CNNs), multiple stacked bidirectional long short-term memory networks (Bi-LSTM), and an attention mechanism to achieve efficient and precise wind power forecasting [9].

In the photovoltaic power generation domain, the types of forecasting models are fundamentally the same as those used for wind power. Monteiro C. et al. in [10] applied data-mining techniques on historical cases to generate power forecasts along with their associated uncertainties or probabilities. David M. et al. in [11] employed a hybrid of two standard econometric linear models (ARMA and GARCH) to produce probabilistic solar-irradiance forecasts. Du Plessis A. et al. in [12] utilized feedforward neural networks, long short-term memory (LSTM) networks, and gated recurrent unit (GRU) recurrent neural networks for PV forecasting, demonstrating the feasibility of deep-learning

approaches in this field.

In 2024, Liu Z. et al. introduced a novel neural-network architecture called Kolmogorov–Arnold Networks (KAN) [13]. Inspired by the Kolmogorov–Arnold representation theorem (detailed in Section 3), they generalised the theorem to arbitrary network widths and depths and on this basis proposed the KAN architecture. Unlike multilayer perceptrons (MLPs), which employ fixed activation functions at their nodes ("neurons"), KAN use learnable activation functions on their edges ("weights")—specifically, B-spline functions. KAN eliminate all linear weight parameters, replacing each one with a univariate spline function, thereby achieving superior accuracy and interpretability compared to MLPs. The original authors later released an enhanced version, KAN 2.0, which improves performance while retaining the core architecture [14]. The comparison among these three model architectures is shown in Figure 1. Accordingly, this work introduces the KAN architecture into the renewable-energy forecasting domain to realize improved predictive performance.



(a) MLP (b) KAN (c) KAN2.0

**Figure 1**. Comparison of the network architectures of MLP, KAN, and KAN 2.0.

The remainder of this paper is organized as follows: Section 2 reviews the latest research advances on KAN; Section 3 presents the theoretical foundations of KAN; and Section 4 describes the experiments conducted in this study along with an analysis of their results.

## 2. Related Work

Since its introduction, KAN has inspired extensive research and applications, leading to numerous variant models that demonstrate its suitability and superiority in areas such as data fitting, complex function learning, high-dimensional data or time-series processing, partial differential equation solving, graph-structured data handling, and hyperspectral image classification [15]. Variants integrating KAN with other architectures—chiefly convolutional, recurrent, and transformer-based models—highlight its versatility when fused with existing neural networks for specialized tasks [16]. Although targeted studies in power systems are still scarce, renewable-energy forecasting fundamentally involves processing and predicting time-series data. Consequently, the following focuses on KAN's advances in time-series analysis:

Vaca-Rubio C. J. et al. demonstrated that, in a real-world satellite-traffic forecasting task, KAN outperforms traditional multilayer perceptrons (MLPs), delivering more accurate predictions with fewer learnable parameters [17]. Genet R. & Inzirillo H. proposed TKAN, a hybrid architecture combining KAN and LSTM strengths, achieving higher accuracy and efficiency in multi-step forecasting tasks [18]. Xu K. et al. introduced two KAN-based variants for time-series processing, emphasizing enhancements in model interpretability to showcase KAN's potential in predictive analytics [19]. Zhou Q. et al. applied KAN to time-series anomaly detection (KAN-AD), converting black-box methods into schemes that learn weights prior to univariate functions, thereby boosting both effectiveness and efficiency [20]. Han X. et al. incorporated a mixture-of-experts structure into KAN to propose a multi-layer mixed KAN network (MMK), demonstrating KAN's efficacy for multivariate time-series forecasting [21].

These studies collectively confirm KAN's feasibility and effectiveness for time-series forecasting, providing strong justification for introducing KAN into renewable-energy prediction tasks in this work.

## 3. KAN

### 3.1. Kolmogorov–Arnold representation theorem

In 1957, the Russian mathematician Andrey Kolmogorov proposed a revolutionary theorem: any continuous multivariate function can be represented as a finite superposition of continuous univariate functions, thereby decomposing complex multivariate mappings into simple one-dimensional components and greatly advancing approximation theory and numerical analysis [22]; subsequently, in 1958, Vladimir Arnold further generalized and confirmed the theorem's universality in higher-dimensional spaces, thus solidifying its theoretical foundation and providing powerful tools for machine learning, data fitting, and partial differential equations [23]. The content of the theorem can equivalently be written in the following form:

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

The principal challenge in applying the Kolmogorov–Arnold representation theorem directly to deep learning is that the theorem inherently involves only two layers of nonlinearity and a limited number of terms (2n + 1). Consequently, the KAN authors argue that, since the majority of functions encountered in practical applications are smooth, it is both natural and necessary to extend the KA representation to networks of arbitrary width and depth to

satisfy modern neural‐network architectural requirements. Extensive empirical studies to date have confirmed the feasibility and efficacy of this generalization.

### 3.2. KAN

The first step in extending the KA representation theorem is to revisit the theorem itself. The KAN authors link the seemingly distinct "external" functions $\Phi_q$ and "internal" functions $\phi_{q,p}$, and the bridge between them is the KAN layer. Specifically, the m-th KAN layer, with $n_m$ input dimensions and $n_{m+1}$ output dimensions, maps an input vector $\mathbf{x}_m \in \mathbb{R}^{n_m}$ to an output vector $\mathbf{x}_{m+1} \in \mathbb{R}^{n_{m+1}}$. By viewing each KAN layer as a means to connect inputs and outputs of arbitrary dimensionality, the KA theorem can thus be generalized to networks of arbitrary depth.

$$\mathbf{x}_{m+1} = \begin{pmatrix} \phi_{m,1,1}(\cdot) & \phi_{m,2,1}(\cdot) & \cdots & \phi_{m,n_m,1}(\cdot) \\ \phi_{m,1,2}(\cdot) & \phi_{m,2,2}(\cdot) & \cdots & \phi_{m,n_m,2}(\cdot) \\ \vdots & \vdots & & \vdots \\ \phi_{m,1,n_{m+1}}(\cdot) & \phi_{m,2,n_{m+1}}(\cdot) & \cdots & \phi_{m,n_m,n_{m+1}}(\cdot) \end{pmatrix} \mathbf{x}_m \quad (2)$$

Connecting the individual KAN layers into a full network:

$$\mathrm{KAN}(\mathbf{x}) = (\boldsymbol{\Phi}_m \circ \cdots \circ \boldsymbol{\Phi}_m \circ \boldsymbol{\Phi}_1)\mathbf{x} \quad (3)$$

Since the KA theorem expresses multivariate mappings as sums of univariate functions, a KAN realizes this by performing summations over its nodes. In practice, this means that even simple multiplicative interactions between variables must be decomposed into several layers and many nodes. To overcome this limitation, KAN 2.0 introduces explicit multiplication nodes, thereby enhancing the network's efficiency and accuracy in tasks such as function approximation.

### 3.3. Activation functions

One key distinction between KAN and a conventional MLP—and indeed one source of KAN's interpretability—is its design and use of learnable univariate activation functions. KAN introduces residual activation functions, whose core formulation is:

$$\phi(x) = w\big(b(x) + \mathrm{spline}(x)\big) \quad (4)$$

where: $b(x)$ is the base branch, $\mathrm{spline}(x)$ is the spline branch, $w$ is a global scaling coefficient.

Similar to a residual connection, the base branch typically employs the SiLU activation function, which endows the network with a well-behaved, smooth nonlinear response from the very start of training (rather than learning entirely from zero). Moreover, retaining a classic activation function aids convergence when data are scarce or gradients are unstable.

$$b(x) = \frac{x}{1 + e^{-x}} \quad (5)$$

The spline branch is constructed as a linear combination of a set of B-spline basis functions. This component serves as a critical source of KAN's learnability and expressiveness. The reasons for choosing B-spline functions are as follows:

(1) Each basis function $B_i$ is nonzero only over a small interval, which means that updating the coefficient $c_i$ affects only a local region of the activation function.

(2) Cubic splines ensure that $\phi(x)$ is at least twice continuously differentiable, which facilitates subsequent analysis involving second-order derivatives of the model.

(3) The trade-off between fitting capacity and overfitting risk can be controlled simply by adjusting the number of knots—more knots allow for greater flexibility, but also increase the risk of overfitting.

$$\mathrm{spline}(x) = \sum_i^N c_i B_i(x) \quad (6)$$

The coefficient $w$ serves as a convenient means to control the overall amplitude of the activation function. Although, in theory, $w$ could be absorbed into the coefficients of $b(x)$ and $\mathrm{spline}(x)$, a separate scaling factor is still retained intentionally.

In summary, this residual-style activation preserves the classical SiLU branch to ensure training stability, while the learnable B-spline branch endows the model with high expressiveness, enabling the precise approximation of any smooth one-dimensional function. Furthermore, after network training is completed, it is possible to visualize and analyze the spline coefficients, thereby enhancing the interpretability of the internal nonlinear transformations. This serves as the primary source of KAN's interpretability.

### 3.4. KAN in renewable energy forecasting

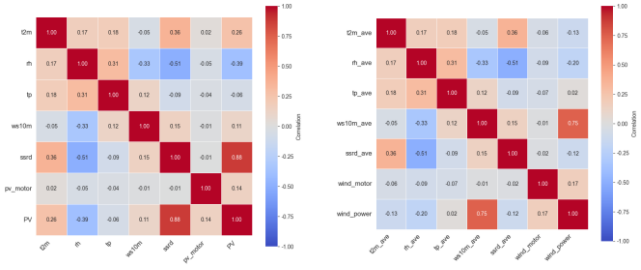
In existing KAN-related research, the targeted application domains primarily include satellite traffic, transportation flow, and weather forecasting. However, through an extensive review of the relevant literature, we have not found any studies that apply KAN to the task of

renewable energy forecasting. This indicates that the application of KAN in this field remains in an exploratory stage and carries significant research value. Compared to traditional power generation methods and other types of time series data, renewable energy generation exhibits higher degrees of uncertainty, randomness, and a certain level of periodicity. These characteristics suggest that the effectiveness of KAN in this context requires empirical validation.

It is also worth noting that, although energy forecasting can be categorized as a time series forecasting task, existing research [21] indicates that KAN does not necessarily achieve optimal performance across all tasks. Therefore, it is essential to explore and design suitable KAN architectures that can effectively address the specific challenges of renewable energy forecasting. The detailed network architecture we adopt will be elaborated in Section 4.

## 4. Experiments

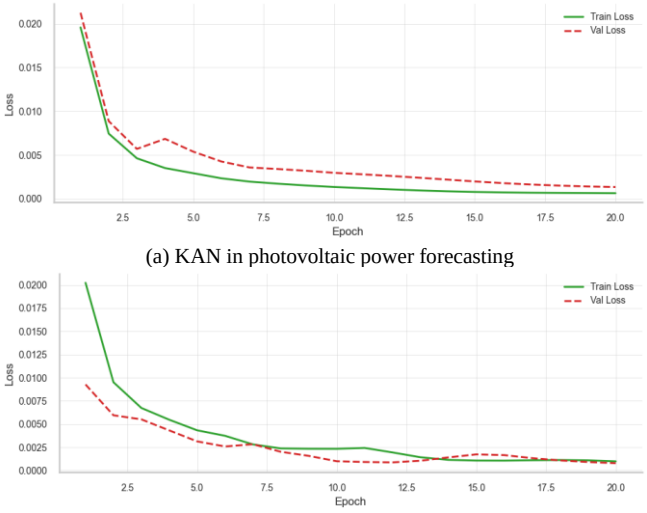
The dataset used in this study originates from a confidential real-world dataset in Shandong Province, China. It comprises hourly time-series data from 2021 to 2023, including wind turbine installed capacity, wind power generation, photovoltaic installed capacity, photovoltaic power generation, and corresponding meteorological data (including temperature, humidity, total precipitation, solar irradiance, and wind speed at 10 meters). Although this dataset is not publicly available, it is regularly maintained and inspected by dedicated personnel, contains no missing values, and can be reasonably assumed to be free from anomalies. Therefore, it is of high quality and well-suited for validating the performance of KAN in the context of renewable energy generation forecasting. The dataset is partitioned into three subsets: training set, validation set, and test set, with respective proportions of 60%, 20%, and 20%. To enhance computational efficiency, data normalization was also applied prior to training.



(a) PV dataset (b) Wind dataset

**Figure 2.** Correlation plot of the dataset

Before initiating model training, and considering that the dataset used is not publicly available, this study conducted a correlation analysis separately on the photovoltaic and wind power components of the dataset along with their associated variables. The results, as illustrated in Figure 2, indicate that wind speed is the primary influencing factor for wind power generation, while solar irradiance is the dominant factor for photovoltaic power generation. These findings are consistent with established physical principles, further confirming the reliability and high quality of the dataset.




(a) KAN in photovoltaic power forecasting

(b) KAN in wind power forecasting

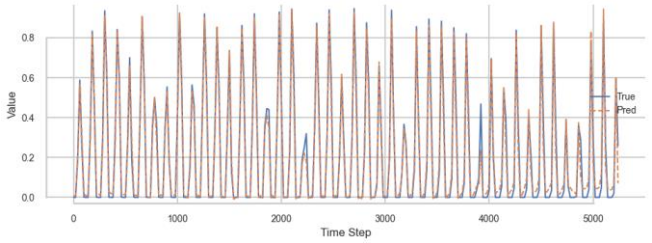
**Figure 3.** Loss function curve during the training process

In this experiment, the baseline models selected for comparison are the Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) networks. On one hand, MLP is a key baseline used by the original authors of KAN, and on the other hand, LSTM has been widely adopted and well-established in the field of energy forecasting. Therefore, using these two models as benchmarks to evaluate the performance of KAN is both justified and reasonable. Each model is tested under different forecasting horizons to simulate varying real-world application needs, including 24-step, 96-step, 168-step, and 336-step predictions. The loss function is set to mean square error (MSE). After training, the models are evaluated on the test set, and both the Mean Absolute Error (MAE) and MSE are calculated for performance assessment.

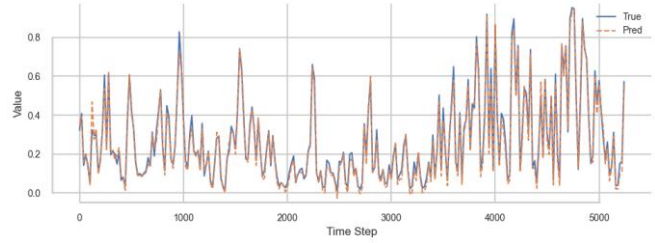
**TABLE 1.** Results summary

| Dataset | Step | Mse($\times 10^3$) | | | Mae ($\times 10^2$) | | |
|---|---|---|---|---|---|---|---|
| | | KAN | LSTM | MLP | KAN | LSTM | MLP |
| Wind | 24 | 2.0 | 2.8 | 7.1 | 2.9 | 3.1 | 5.6 |
| | 96 | 4.8 | 5.4 | 11.1 | 4.1 | 4.8 | 6.9 |
| | 168 | 10.3 | 11.1 | 30.1 | 6.7 | 6.9 | 10.7 |
| | 336 | 24.9 | 23.1 | 44.1 | 7.7 | 7.0 | 13.9 |
| PV | 24 | 1.5 | 2.3 | 6.2 | 2.6 | 2.8 | 5.1 |
| | 96 | 4.3 | 4.9 | 10.6 | 3.4 | 3.9 | 6.4 |
| | 168 | 8.9 | 9.4 | 22.6 | 5.2 | 5.7 | 9.7 |
| | 336 | 15.3 | 15.6 | 43.8 | 6.4 | 6.6 | 12.1 |

Model Configuration Used in the Experiment: For the KAN network, we adopted a structure with four hidden layers, where the input and output dimensions vary depending on the specific forecasting horizon. The MLP model, similar to KAN, also consists of four layers and uses the ReLU function as its activation function. The LSTM model was configured with two hidden layers. Through preliminary experiments, we observed that when trained with the Adam optimizer, the networks' loss function (MSE) generally ceased to decrease after 20 epochs. Consequently, the final experiments were conducted using 20 training epochs.




(a) Photovoltaic power forecasting




(b) Wind power forecasting

**Figure 4.** Visualization of prediction performance on the test set

The training process of KAN is shown in Figure 3. Once the model converged, its performance was evaluated on the test set. The predictive performance comparison of the models on the test set is illustrated in Figure 4 (To safeguard the sensitivity of the dataset, all data presented in the figure have been normalized). It should be noted that due to the large number of data samples in the test set, only a portion of the 96-step samples were selected for visualization to facilitate clearer comparison. A comprehensive comparison of the global prediction performance of each model across different time horizons and for both energy types is summarized in Table 1. It is evident that KAN consistently achieves the best performance in most scenarios. Specifically, for the 96-step and 168-step wind power forecasting tasks, KAN outperformed the second-best model, LSTM, with reductions in MSE of 11.1% and 7.2%, respectively. For photovoltaic power forecasting at certain time steps, KAN achieved MSE reductions of 12.2% and 5.3% compared to LSTM. In the 336-step forecasting task, KAN's performance was approximately on par with that of LSTM. An additional noteworthy observation is that all models performed slightly better on photovoltaic power forecasting than on wind power forecasting. This can be attributed partly to the higher periodicity of photovoltaic generation, and partly to the significantly greater uncertainty inherent in wind power generation. Taken together, these results indicate that KAN not only delivers strong predictive performance in the field of energy forecasting but also retains its advantage of interpretability.

## 5. Conclusions

In this study, KAN has been successfully introduced into the task of renewable energy forecasting, and its strong performance has been demonstrated. Based on a comprehensive review of existing research, we first established the feasibility of our approach. Subsequently, using real-world wind and photovoltaic power generation datasets, we conducted comparisons with MLP and LSTM models, clearly validating the superior performance of KAN in energy forecasting tasks. Currently, the effectiveness of KAN in time series applications is a topic of active discussion, with numerous promising studies emerging. The results presented in this work further support the potential of KAN as a valuable tool for time series forecasting.

## References


[1] Wang, Yun, et al. "Approaches to wind power curve modeling: A review and discussion." Renewable and Sustainable Energy Reviews 116 (2019): 109422.

[2] Wang, Yun, et al. "Sparse heteroscedastic multiple spline regression models for wind turbine power curve modeling." IEEE Transactions on Sustainable Energy 12.1 (2020): 191-201.

[3] Yang, Jian, et al. "Review of control strategy of large horizontal‐axis wind turbines yaw system." Wind Energy 24.2 (2021): 97-115.

[4] Ahmed, Razin, et al. "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization." Renewable and sustainable energy reviews 124 (2020): 109792.

[5] Wang, Yun, et al. "A review of wind speed and wind power forecasting with deep neural networks." Applied energy 304 (2021): 117766.

[6] Guo, Nai-Zhi, et al. "A physics-inspired neural network model for short-term wind power prediction considering wake effects." Energy 261 (2022): 125208.

[7] Kim, Yeojin, and Jin Hur. "An ensemble forecasting model of wind power outputs based on improved statistical approaches." Energies 13.5 (2020): 1071.

[8] Kisvari, Adam, Zi Lin, and Xiaolei Liu. "Wind power forecasting–A data-driven method along with gated

recurrent neural network." Renewable Energy 163 (2021): 1895-1909.
[9] Ma, Zhengjing, and Gang Mei. "A hybrid attention-based deep learning approach for wind power prediction." Applied Energy 323 (2022): 119608.
[10] Monteiro, Claudio, et al. "Short-term power forecasting model for photovoltaic plants based on historical similarity." Energies 6.5 (2013): 2624-2643.
[11] David, Mathieu, et al. "Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models." Solar Energy 133 (2016): 55-72.
[12] Du Plessis, A. A., J. M. Strauss, and A. J. Rix. "Short-term solar power forecasting: Investigating the ability of deep learning models to capture low-level utility-scale Photovoltaic system behaviour." Applied Energy 285 (2021): 116395.
[13] Liu, Ziming, et al. "Kan: Kolmogorov-arnold networks." arXiv preprint arXiv:2404.19756 (2024).
[14] Liu, Ziming, et al. "Kan 2.0: Kolmogorov-arnold networks meet science." arXiv preprint arXiv:2408.10205 (2024).
[15] Ji, Tianrui, Yuntian Hou, and Di Zhang. "A comprehensive survey on Kolmogorov Arnold networks (KAN)." arXiv preprint arXiv:2407.11075 (2024).
[16] Somvanshi, Shriyank, et al. "A survey on kolmogorov-arnold network." arXiv preprint arXiv:2411.06078 (2024).
[17] Vaca-Rubio, Cristian J., et al. "Kolmogorov-arnold networks (kans) for time series analysis." arXiv preprint arXiv:2405.08790 (2024).
[18] Genet, Remi, and Hugo Inzirillo. "Tkan: Temporal kolmogorov-arnold networks." arXiv preprint arXiv:2405.07344 (2024).
[19] Xu, Kunpeng, Lifei Chen, and Shengrui Wang. "Kolmogorov-arnold networks for time series: Bridging predictive power and interpretability." arXiv preprint arXiv:2406.02496 (2024).
[20] Zhou, Quan, et al. "KAN-AD: Time series anomaly detection with Kolmogorov-Arnold networks." arXiv preprint arXiv:2411.00278 (2024).
[21] Han, Xiao, et al. "Kan4tsf: Are kan and kan-based models effective for time series forecasting?." arXiv preprint arXiv:2408.11306 (2024).
[22] Kolmogorov, Andreĭ Nikolaevich. On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables. American Mathematical Society, 1961.
[23] Arnold, Vladimir I. "On functions of three variables." Collected Works: Representations of Functions, Celestial Mechanics and KAM Theory, 1957–1965 (2009): 5-8.