Systems & Control Transactions

Original Research Article
Peer Reviewed Conference Proceeding
ESCAPE 36 - European Symposium on Computer Aided Process Engineering
Sheffield, UK. 21-24 June 2026

PSE PRESS

Deep Kernel Learning with Kolmogorov–Arnold Networks for Bayesian Optimization

Zhongtao Shang[a], Zhihong Yuan[b], Lifeng Zhang[c]*, Yiyang Dai[a]*

[a] School of Chemical Engineering, Sichuan University, Chengdu, 610065, P. R. China.
[b] Department of Chemical Engineering, Tsinghua University, Beijing, 100084, China.
[c] Department of Chemical Engineering, The Sargent Centre for Process Systems Engineering, Imperial College London, London, SW7 2AZ, United Kingdom
* Corresponding Author: lifeng.zhang@imperial.ac.uk, daiyy@scu.edu.cn

ABSTRACT

Deep Kernel Learning (DKL) has emerged as a powerful framework for Bayesian Optimization (BO), via combining expressive representation learning models with typical Gaussian Processes (GPs) surrogate models. However, conventional DKL typically relies on weight-based feature extractors (e.g., multilayer perceptrons (MLPs)), which often lack interpretability and may suffer from overfitting under data scarcity or training instability, potentially leading to a degraded uncertainty quantification in GP models. Grounded in the Kolmogorov-Arnold representation theorem, this paper proposes a novel DKL-KAN framework that employs Kolmogorov-Arnold Networks (KANs) as adaptive feature extractors, formulating a DKL-KAN surrogate model. Unlike MLPs, the KANs learn data-driven univariate functions, yielding more sample-efficient and stable representations for regression under limited data regimes. Followed by the GP, the DKL-KAN facilitates end-to-end learning of expressive latent representations while maintaining robust uncertainty quantification. The proposed DKL-KAN framework is further validated with the Williams-Otto (W-O) process benchmarks. Compared to classical GPs and DKL-MLPs, the DKL-KANs consistently exhibit superior predictive accuracy and uncertainty calibration even under high-dimensional and scarce dataset. The results show that the KANs generate more structured and informative latent embeddings, thus enhancing the ability to capture complex physical nonlinearities in sparsely sampled edge regions of the design space where traditional models often falter. Moreover, the DKL-KAN is embedded into the BO and yields the efficient optimization performance on the W–O process benchmark, achieving faster early-stage improvement and converging to higher objective values. These results demonstrate the potential of DKL-KAN surrogates for accelerating BO in chemical engineering processes.

Keywords: Deep Kernel Learning, Kolmogorov-Arnold Network, Bayesian Optimization, Process Optimization

1. INTRODUCTION

Black-box optimization (BBO) has been a fundamental problem in many scientific and industrial applications, including process optimization[1], experimental design[2], drug discovery[3] and model hyperparameter tuning[4], where the objective function is expensive to evaluate and lacks accessible analytical expressions or gradient information. In the past decades, Bayesian Optimization (BO) has emerged as a premier framework in BBO, renowned for its ability to search global optima by balancing exploration and exploitation[5].

In BO, the crucial step is to develop a surrogate model to offer not only the mean predictions but also the uncertainty quantification for evaluating the acquisition function. Therefore, the Gaussian Process (GP) has been widely used in BO owing to its principled probabilistic formulation and kernel-based covariance structure. However, when applied to high-dimensional datasets, the GPs often suffer from significant computational costs for constructing the kernel matrix and kernel degeneracy, which may degrade the model performance. To resolve these limitations, Deep Kernel Learning (DKL) has been proposed to automatically learn the low-dimensional

latent embeddings and then feeding them to the GP models[6–8]. In practice, the MLPs are commonly used in most existing DKL approaches for their ability to learn flexible nonlinear feature representations from high-dimensional inputs. Singh et al.[9] proposed MLP-based DKL surrogate to convert multiple molecular representations to the latent space embedding and achieve accurate reaction yield predictions. Andrea et al.[10] developed a closed-loop DKL-based BO framework for porous lithium-ion cathode microstructure design, where a Generative Adversarial Network(GAN) generator is used to define the latent design space. Nevertheless, the limited interpretability, susceptibility to overfitting under scarce data, and poor training stability of MLPs often restrict their applicability in complex process optimization problems. Furthermore, according to prior studies, an inappropriate and over-parameterized feature extractor in DKL can lead to unreliable uncertainty estimation in BO[7].

Inspired by the Kolmogorov–Arnold representation theorem, Kolmogorov–Arnold Networks (KANs)[11], have recently emerged as a promising alternative approach to traditional MLPs. Unlike MLPs that rely on predefined activation functions to capture the nonlinearity, KANs employ learnable univariate functions as adaptive activations, enabling them to capture complex nonlinear relationships with enhanced interpretability and data efficiency. Numerous studies have demonstrated the great potential of KANs in modeling fluid dynamics, as well as in executing time-series classification and differential equation solvers[12, 13]. By providing more interpretable functional representations, KANs have shown more attraction for scientific discovery and mechanism-oriented analysis compared to MLPs[14].

Motivated by these observations, this work proposes a DKL-KAN model, in which a KAN serves as the feature extractor and is followed by a GP model within the DKL framework. The DKL-KAN model simultaneously learns the expressive latent representations and quantify the uncertainty, revealing good capture of data distribution and accurate predictions. The proposed DKL-KAN model is further integrated into the BO framework and applied to the benchmark Williams–Otto (W-O) process optimization problem. Compared to standard GPs (std-GPs) and DKL-MLP, the DKL-KAN model demonstrates a competitive performance, maintaining superior predictive accuracy, uncertainty calibration, and optimization efficiency even under high-dimensional and data-scarce regimes. Moreover, the DKL-KAN framework highlights improved interpretability and structural insight into the underlying process. The proposed DKL-KAN model offers new insights for addressing complex chemical process optimization problems.

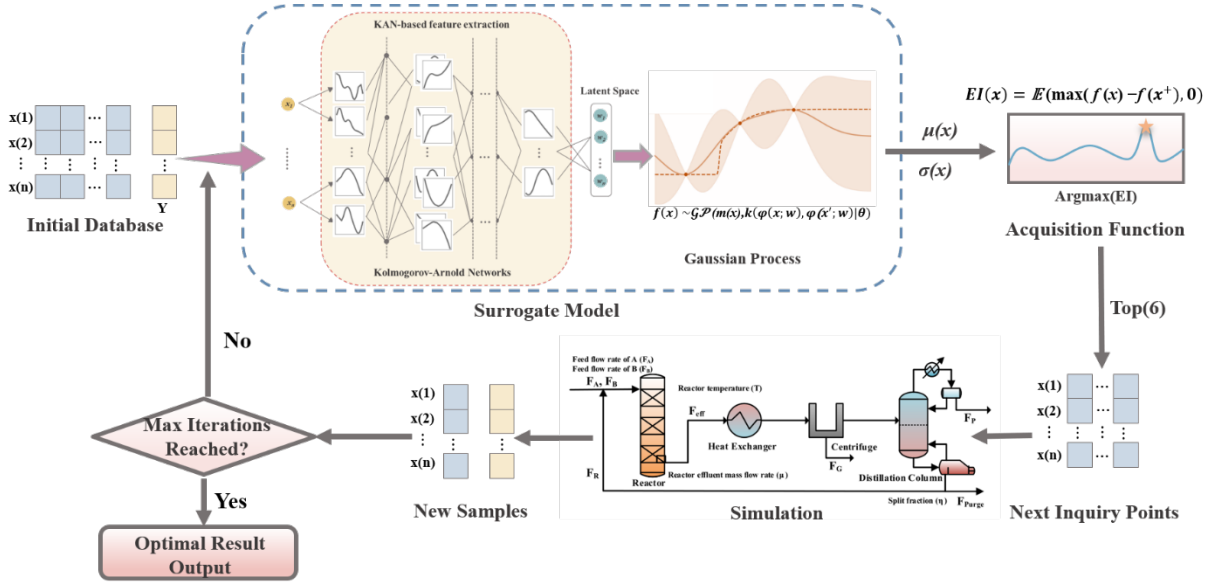

Figure 1. Schematic illustration of DKL-KAN framework for BO.

2. METHODLOGY

2.1 Optimization framework

In this work, a complex process optimization problem is considered to obtain the optimal operations under limited evaluation budget, which can be formulated as follows:

$$\max\ y = f(x) \tag{1}$$

$$s.t.\, x^{lb} \le x \le x^{ub}$$

$$x \subseteq R^n$$

Where y denotes a scalar metric to be optimized, such as

annual profit or process yields. x represents continuous decision variables, such as temperature, flow rate, and pressure. To solve the problems, the overall optimization framework which combines DKL-KAN and BO can be illustrated in Figure 1. The framework is structured as follows:

1. A small-batch dataset is generated with Sobol' sequence[15] from the design space and evaluated to get the scaler metric via process simulations.

2.With the existing dataset, the DKL-KAN surrogate model is developed by combining a KAN model as a feature extractor and a GP model which maps the normalized latent embeddings to predictive means and covariances in an end-to-end training manner.

3.Using the DKL–KAN surrogate, the expected improvement (EI) values, also known as the acquisition function, are evaluated over the candidate set. A next batch of sampling points are determined by maximizing the EI.

4.The suggested candidates from EI are subsequently evaluated in the simulation environment or retrieved from the dataset to obtain new observations, which are appended to the training set for updating the surrogate model in the next iteration.

5.The procedure terminates when the maximum number of validation iterations is reached, and the optimal solution identified so far is reported; otherwise, the workflow returns to Step 2 to update the surrogate model and continues iteratively.

2.2 Kolmogorov-Arnold Networks

KANs are theoretically grounded in the Kolmogorov-Arnold representation theorem[11] which asserts that any multivariate continuous function $f: [0,1]^n \to \mathbb{R}$ can be decomposed as a combination of univariate functions:

$$f(\mathbf{x}) = f(x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} \phi_q \left(\sum_{p=1}^{n} \varphi_{q,p}(x_p) \right) \tag{2}$$

where $\varphi_{q,p}$ and ϕ_q denote univariate continuous functions. While the classical theorem restricts the architecture to a fixed depth, modern KANs generalize this into a deep stacked structure. For a network with $|L|$ layers, the forward pass is represented as a composition of functional layers:

$$\mathrm{KAN}(\mathbf{x}) = \left(\phi_{|L|-1} \circ \phi_{|L|-2} \circ \ldots \circ \phi_1\right)\mathbf{x} \tag{3}$$

where L refers the set of layers and each layer ϕ_l consists of a set of trainable 1-D activation functions $\{\varphi_{l,j}\}$. In practice, each univariate activation $\varphi_{l,j}(x)$ is parameterized as a combination of basis function $b_{l,j}(x)$ and a B-spline function:

$$\varphi_{l,j}(x) = \omega_{l,j} \left[b_{l,j}(x) + \sum_{m} c_{l,j,m} B_{l,j,m}(x) \right] \quad \forall l \in L, j \in J^i \tag{4}$$

Where $m = 1, \ldots, M$ indexs B-spline basis functions, and $B_{l,j,m}(x)$ denotes the m^{th} B-spline basis defined on the corresponding grid. The spline coefficients $c_{l,j,m}$ and grid configurations in KANs are fully learnable, allowing the network to adaptively refine its function approximation capabilities.

2.3 DKL-KAN Surrogate Model

The DKL-KAN surrogate model for BO is formulated with KAN and GP where the GP converts the latent embeddings to predictive means and uncertainty quantification. Generally, the GP regression is defined as follows to obtain the black-box function $f(\mathbf{x})$[16]:

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right) \tag{5}$$

where $m(\mathbf{x})$ and $k(\mathbf{x}, \mathbf{x}')$ denote the mean function and kernel function. Given observations $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, GP posterior yields a predictive distribution at any candidate $\mathbf{x}_i$, whose posterior mean $\eta(\mathbf{x}|\mathcal{D})$ serves as the prediction and posterior variance $\sigma^2(\mathbf{x}|\mathcal{D})$ quantifies epistemic uncertainty. This probabilistic output is directly leveraged by acquisition functions to guide sampling in BO.

Besides, conventional GP surrogates can be limited in practical applications. Exact GP inference scales cubically with the number of evaluations due to kernel matrix computation. Moreover, the representational capacity of GP is strongly determined by the choice of kernel, standard stationary kernels may struggle to capture complex nonstationarity, sharp local variations, or high-dimensional structure without substantial kernel engineering. DKL mitigates these drawbacks by learning a task-adaptive representation prior to GP modeling. Given a KAN feature map $g(\mathbf{x}|\beta)$arameterized by β, and $k(\mathbf{x}, \mathbf{x}'|\alpha)$ representing the standard kernel parameterized by α, the kernel function can be formulated as follows[17]:

$$k_{DKL-KAN}(\mathbf{x}, \mathbf{x}') = k_{base}(g(\mathbf{x}_i; \beta), g(\mathbf{x}_i', \beta)|\alpha, \beta) \tag{6}$$

This construction preserves GP-style uncertainty quantification while improving expressiveness through data-driven feature learning, enhancing the model capacity for real world applications.

Following the DKL-KAN formulation introduced above, the surrogate is trained by jointly learning β and α via maximizing the log marginal likelihood $\mathcal{L}$ [18]. Let $\mathbf{y} \in \mathbb{R}^n$ denote the observations at inputs $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{n}$, and let $\mathbf{K} \in \mathbb{R}^{n \times n}$ be the covariance matrix with entries $[\mathbf{K}]_{ij} = k_{\mathrm{DKL-KAN}}(\mathbf{x}_i, \mathbf{x}_j)$. Assuming Gaussian noise with variance σ_n^2, the $\mathcal{L}$ is given by:

$$\mathcal{L}(\alpha, \beta) = -\frac{1}{2}\mathbf{y}^{\top}(\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K} + \sigma_n^2\mathbf{I}| - \frac{n}{2}\log(2\pi) \tag{7}$$

Then the surrogate model is obtained via minimizing the negative log marginal likelihood (NLL):

$$\alpha^{\star}, \beta^{\star} = \arg\min_{\alpha,\beta} -\mathcal{L} \tag{8}$$

Compared to the commonly used DKL-MLP surrogate which is usually an opaque, black-box feature extractor, the proposed DKL-KAN provides a more structured and inspectable embedding with improved data efficiency and uncertainty calibration, particularly for high-dimensional, data-scarce chemical process optimization.

2.4 Acquisition Function

A variety of acquisition functions have been proposed for BO, including probability of improvement (PI), EI, and upper confidence bound (UCB), which differ in how to trade off exploration and exploitation[19]. Among them, EI is particularly attractive because it directly quantifies the expected gain over the current best observation[20]. Given the surrogate predictive distribution $y \sim \mathcal{N}(\mu(\mathbf{x}), \sigma^2(\mathbf{x}))$, EI at a new point $\mathbf{x}$ can be written in closed-form as:

$$\mathrm{EI}(\mathbf{x}) = \big(\mu(\mathbf{x}) - f(\mathbf{x}^*)\big)\Phi(Z) + \sigma(\mathbf{x})\phi(Z) \tag{9}$$

where $f(\mathbf{x}^*)$ denotes the best objective value observed so far, Φ and ϕ denote the cumulative distribution function (CDF) and the probability density function (PDF) of the standard normal distribution, respectively. Z is a random variable defined as:

$$Z(\mathbf{x}) = \frac{\mu(\mathbf{x}) - f(\mathbf{x}^*)}{\sigma(\mathbf{x})} \tag{10}$$

In the proposed framework, $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ are provided from the DKL-KAN. At each iteration, EI is evaluated over the possible candidate set and a batch of the six points with the largest EI values is selected for querying. The resulting objective evaluations are then added to the training set to update the surrogate, and the procedure is repeated until the maximum iterations is reached.

3.CASE STUDY

3.1 Problem description

The proposed method is then applied to the W-O process optimization problem as case study. For a brief introduction, the W-O process is consist of a continuously stirred tank reactor (CSTR), a decanter, a distillation column (separator) and a splitter[21], as shown in Figure 2. The decision variables and respective bounds for W-O process can be defined in Table 1. F_{A} and F_{B} denote the feed flow rates of reactants A and B, T is the operating temperature of reactor, μ represents the total mass flow rate of the reactor effluent, and η specifies the split fraction.

Under a series of reactions, the objective for W-O process is to maximize the Return On Investment (ROI) as follows:

$$ROI = \frac{Annual\ Net\ Profit}{Total\ Capital\ Investment} \tag{11}$$

For the details of W-O process, including reaction system, mass balance constraints and reaction kinetic equations, the readers are referred to the relevant literature[21].

All computations are conducted on a workstation equipped with an Intel(R) Core(TM) i9-14900KF CPU (3.20 GHz) and 64 GB RAM. The W-O is simulated via *Python* and the KAN is constructed using the *pykan*, while GP is constructed with *GPyTorch*.

Table 1. Decision variable bounds of the optimization process.

Variables	Unit	Bounds
F_A	klb/h	[2.0, 40]
F_B	klb/h	[2.0, 40]
T	K	[520, 700]
μ	klb/h	[10, 200]
η	-	[0, 0.9]

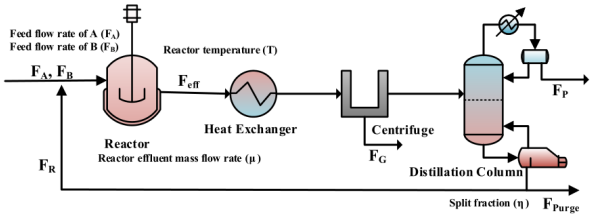

Figure 2. Schematic illustration of W-O process.

3.2 Performance of DKL-KAN

To assess the predictive capability of the proposed DKL–KAN surrogate, multiple datasets with different sample sizes are generated. For each dataset, samples were randomly split into training and test sets with an 8:2 ratio. Five models are developed for comparison, namely, std-GP, DKL–MLP, DKL–KAN, std-MLP, and std-KAN, where *std* denotes the standard model. Hyperparameter tuning is conducted by Optuna[22] to ensure a fair comparison under consistent hyperparameter optimization. The network architectures are eventually fixed as KAN: [5×4×2] and MLP: [5×64×32×2].

The R-squared (R^2) values with the dataset of 500 points are summarized in Figure 3 (a). Obviously, DKL-based surrogates improve generalization over std-GP, and DKL–KAN delivers superior performance. Worse still, std-GP is highly kernel-sensitive with R^2 ranging from 0.609±0.083(linear) to 0.794±0.068 (Radial Basis Function, RBF), indicating that the kernels cannot adequately capture the complex underlying feature structure. DKL–KAN maintains high R^2 across different kernels and data sizes, showing consistent performance from 0.919± 0.045, 0.935±0.036 and 0.916±0.056 for linear, RBF and

Matérn-2.5 kernels, respectively. This indicates that introducing KAN for DKL can significantly improve the model capacity regardless the kernels. In contrast, the baselines (std-GP, std-MLP and std-KAN) exhibit larger performance gaps and variability, whereas std-KAN attains relatively high R^2 with 0.926±0.105, but lacks the predictive uncertainty required. This superiority is further reflected in the root mean squared error (RMSE) metrics trends across dataset sizes (Figure 3 (b)), DKL–KAN yields the lowest test RMSE, with its advantage being most evident in the low-data regime and persisting as the training set grows. This enhanced performance is attributed to the integration of the KANs within the deep kernel architecture, which enables more expressive nonlinear feature mapping and superior uncertainty quantification compared to fixed-basis surrogates.

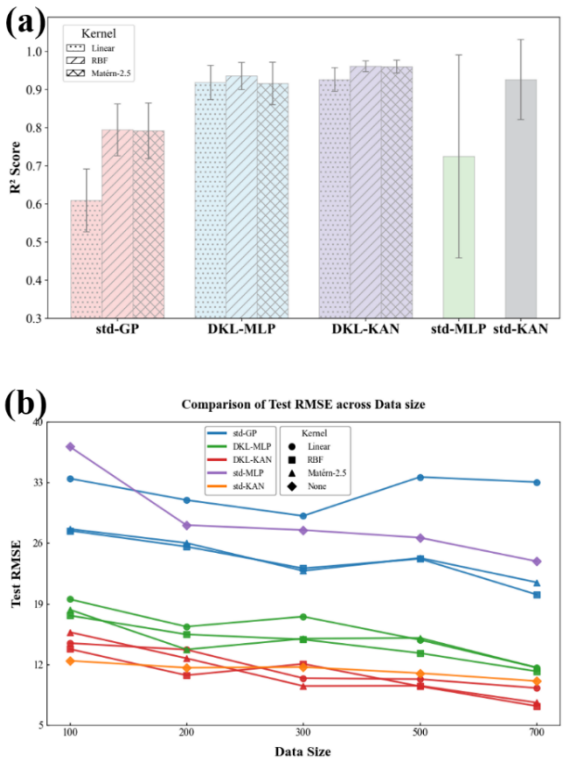

Figure 3. (a) R^2values at dataset of 500 points. (b) Test set RMSEs on different sample sizes.

3.3 Latent space embedding analysis

To gain deeper insights into the behavior of the models, the latent space embeddings of DKL–KAN model is analyzed and compared with the DKL–MLP and std-GP models. For the DKL-driven models, the two-dimensional embeddings from the final feature mapping layer are directly extracted, while for the std-GP, UMAP dimensionality reduction is applied[23]. Subsequently, k-means clustering[24] is performed to obtain four distinct clusters. As shown in Figure 4, the introduction of DKL significantly enhances the distribution of the latent space compared to the std-GP baseline, with clusters becoming more compact and well-separated[9]. The ROI color map reveals a clearer spatial distribution, indicating improvement by introducing DKL-driven surrogates. A closer examination of the cluster characteristics, the highest-ROI cluster is associated with higher reactor temperatures, larger outlet flow rates, and smaller split fractions, while lower-ROI clusters correspond to different combinations of these decision variables.

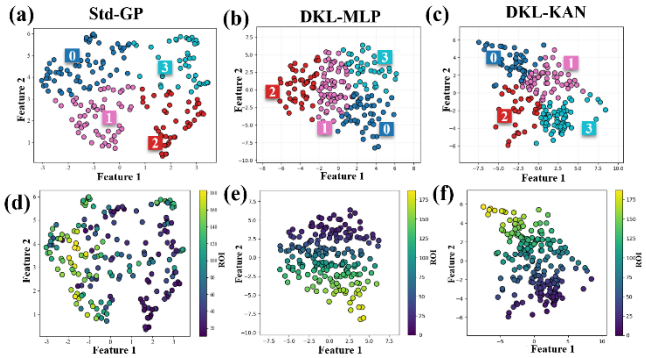

Figure 4. Analysis of the latent space of the (a) std-GP, (b) DKL-MLP and (c) DKL-KAN. The colors represent different clusters identified using k-means clustering. The distribution of ROI in each cluster formed from the (d) std-GP, (e) DKL-MLP and (f) DKL-KAN.

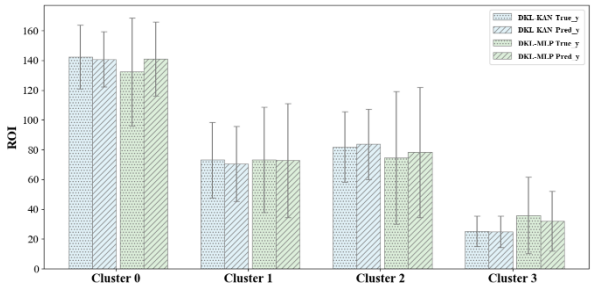

Figure 5. Comparison of ROI Prediction Consistency Across Clusters for DKL–MLP, DKL–KAN.

Figure 5 further illustrates the model predictions within each cluster. DKL–KAN demonstrates consistent ROI predictions across all clusters, performing well even in both high- and low-ROI regions, thus highlighting its ability to handle complex nonlinearities effectively. In contrast, DKL-MLP shows more significant fluctuations in ROI predictions across clusters, alternating between overestimations and underestimations, especially in the data-sparse, high-nonlinearity extreme regions, reflecting a fundamental limitation. While the integration of GP priors offers uncertainty quantification, the MLP feature extractor struggles to capture the highly nonlinear, multimodal nature of complex objective functions. The inherent smoothness of the MLP architecture, combined with its inability to adapt to sharp variations or multiple peaks,

leads to suboptimal performance, especially in data-scarce regions with high-dimensional nonlinearity. These results underscore the ability of DKL–KAN to effectively capture the underlying structure of the W–O process, enhancing both the separability of the learned feature embeddings and the accuracy of the predictions.

3.4 Optimisation performance

The DKL-KAN is further integrated with BO as the surrogate model for W-O process to validate the optimization efficiency[25]. The total optimization iteration is set to 20, with performance reported as the mean and standard deviation across five independent trials with randomized initializations, illustrated in Figure 6. Computational results from std-GP and DKL-MLP are also provided for comparison. Both DKL-MLP and DKL-KAN surrogates outperform the std-GP in terms of optimization efficiency, consistent with prior work showing that deep kernel learning can improve surrogate accuracy and thus accelerate BO convergence by capturing informative feature representations than shallow kernels alone. Notably, the proposed DKL–KAN model exhibits the most efficient performance, achieving faster improvement in the early iterations and converging to higher objective values. This superior performance is primarily attributable to the DKL–KAN surrogate, whose KAN feature extractor learns more structrued latent embeddings than MLPs and thus the GP kernel can better distinguish the datapoints and captures the highly nonlinearity. Overall, these results demonstrate that the proposed DKL–KAN model provides both excellent predictive accuracy and reliable uncertainty quantification, making it a promising alternative method for process optimization tasks where high-dimensional, nonlinear relationships are prevalent.

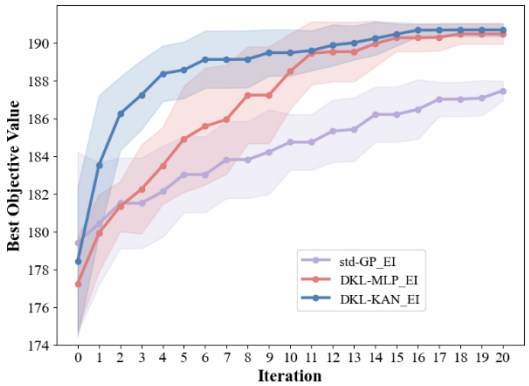

Figure 6. Comparison of optimization performance under EI acquisition function with std-GP, DKL-MLP and DKL-KAN surrogate models.

4. CONCLUSIONS AND FUTURE WORK

This work develops a novel DKL surrogate model leveraging the KAN as an adaptive feature extractor. Systematically validated on the W–O process benchmark, the proposed DKL-KAN model demonstrates superior predictive performance and more reliable uncertainty quantification compared to standard GPs and conventional DKL-MLP. Latent space analysis further elucidates that the KAN facilitates a more structured embedding distribution, maintaining high generalization and accurate uncertainty characterization even in sparse boundary regions of the sample space. When integrated with BO, the approach identifies superior optimum with significantly accelerated convergence in early iterations. This can be attributed to the more informative embeddings from KAN which better represent the highly nonlinear and strongly coupled W–O response surface. These findings underscore the significant potential of DKL-KAN surrogates as a promising method for complex chemical engineering process optimization. Future work can be done to incorporate advanced algorithms within the latent space to enhance the robustness and efficiency for global optimization exploration.

ACKNOWLEDGEMENTS

The authors are grateful for the support of the National Natural Science Foundation of China (22538010).

REFERENCES

1. Zhou J, Shi T, Ren J, He C. Accelerating operation optimization of complex chemical processes: a novel framework integrating artificial neural network and mixed-integer linear programming. Chemical Engineering Journal 481:148421 (2024). https://doi.org/10.1016/j.cej.2023.148421
2. Shields BJ, Stevens J, Li J, Parasram M, Damani F, Alvarado JIM, Janey JM, Adams RP, Doyle AG. Bayesian reaction optimization as a tool for chemical synthesis. Nature 590:89-96 (2021). https://doi.org/10.1038/s41586-021-03213-y
3. Li H, Hao J, Qiao S. Ai?driven electrolyte additive selection to boost aqueous zn?ion batteries stability. Advanced Materials 36: (2024). https://doi.org/10.1002/adma.202411991
4. S. Falkner, A. Klein, F. Hutter, BOHB: Robust and Efficient Hyperparameter Optimization at Scale, (2018). https://doi.org/10.48550/arXiv.1807.01774
5. Winz J, Fromme F, Engell S. Bayesian optimization of gray-box process models using a modified upper confidence bound acquisition function. Computers & Chemical Engineering 194:108976 (2025). https://doi.org/10.1016/j.compchemeng.2024.108976
6. A.G. Wilson, Z. Hu, R. Salakhutdinov, E.P. Xing, Deep Kernel Learning, in: Proceedings of the 19th

International Conference on Artificial Intelligence and Statistics, PMLR, 2016: pp. 370–378. https://proceedings.mlr.press/v51/wilson16.html (accessed October 10, 2025).
7. I. Achituve, G. Chechik, E. Fetaya, Guided Deep Kernel Learning, (2023). https://doi.org/10.48550/arXiv.2302.09574
8. Botteghi N, Guo M, Brune C. Deep kernel learning of dynamical models from high-dimensional noisy data. Sci Rep 12: (2022). https://doi.org/10.1038/s41598-022-25362-4
9. Singh S, Hernández-Lobato JM. Deep kernel learning for reaction outcome prediction and optimization. Commun Chem 7: (2024). https://doi.org/10.1038/s42004-024-01219-x
10. Gayon-Lombardo A, del Rio-Chanona EA, Pino-Muñoz CA, Brandon NP. Deep kernel bayesian optimisation for closed-loop electrode microstructure design with user-defined properties. Energy and AI 22:100608 (2025). https://doi.org/10.1016/j.egyai.2025.100608
11. Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T.Y. Hou, M. Tegmark, KAN: Kolmogorov-Arnold Networks, (2025). https://doi.org/10.48550/arXiv.2404.19756
12. S.A. Faroughi, F. Mostajeran, A.H. Mashhadzadeh, S. Faroughi, Scientific Machine Learning with Kolmogorov-Arnold Networks, (2025). https://doi.org/10.48550/arXiv.2507.22959
13. Wang Y, Sun J, Bai J, Anitescu C, Eshaghi MS, Zhuang X, Rabczuk T, Liu Y. Kolmogorov–arnold-informed neural network: a physics-informed deep learning framework for solving forward and inverse problems based on kolmogorov–arnold networks. Computer Methods in Applied Mechanics and Engineering 433:117518 (2025). https://doi.org/10.1016/j.cma.2024.117518
14. Zeng Y, Cao W, Zuo Y, Peng T, Hou Y, Miao L, Wang Z, Shi J. Accelerating the discovery of materials with expected thermal conductivity via a synergistic strategy of DFT and interpretable deep learning. Mater. Futures 4:045602 (2025). https://doi.org/10.1088/2752-5724/ae08d0
15. E. Atanassov, S. Ivanovska, On the Use of Sobol' Sequence for High Dimensional Simulation, in: D. Groen, C. de Mulatier, M. Paszynski, V.V. Krzhizhanovskaya, J.J. Dongarra, P.M.A. Sloot (Eds.), Computational Science – ICCS 2022, Springer International Publishing, Cham, 2022: pp. 646–652. https://doi.org/10.1007/978-3-031-08760-8_53
16. Richardson RR, Osborne MA, Howey DA. Gaussian process regression for forecasting battery state of health. Journal of Power Sources 357:209-219 (2017). https://doi.org/10.1016/j.jpowsour.2017.05.004
17. S.W. Ober, C.E. Rasmussen, M. van der Wilk, The Promises and Pitfalls of Deep Kernel Learning, (2021). https://doi.org/10.48550/arXiv.2102.12108
18. C.E. Rasmussen, C.K.I. Williams, Gaussian processes for machine learning, 3. print, MIT Press, Cambridge, Mass., 2008.
19. E. Brochu, V.M. Cora, N. de Freitas, A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning, (2010). https://doi.org/10.48550/arXiv.1012.2599
20. S. Ament, S. Daulton, D. Eriksson, M. Balandat, E. Bakshy, Unexpected Improvements to Expected Improvement for Bayesian Optimization, (2025). https://doi.org/10.48550/arXiv.2310.20708
21. X. Jin, K.A. High, Decision Making for a Sustainable Chemical Process, (n.d.).
22. T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A Next-generation Hyperparameter Optimization Framework, (n.d.).
23. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, (n.d.).
24. Ikotun AM, Ezugwu AE, Abualigah L, Abuhaija B, Heming J. K-means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data. Information Sciences 622:178-210 (2023). https://doi.org/10.1016/j.ins.2022.11.139
25. L.N. Sridhar, Multiobjective Nonlinear Model Predictive Control of the Williams Otto Process, Chemical Engineering 1 (2024).