Energy and AI 20 (2025) 100473

Contents lists available at ScienceDirect

Energy and AI

journal homepage: www.sciencedirect.com/journal/energy-and-ai

# Comparison of Kolmogorov–Arnold Networks and Multi-Layer Perceptron for modelling and optimisation analysis of energy systems

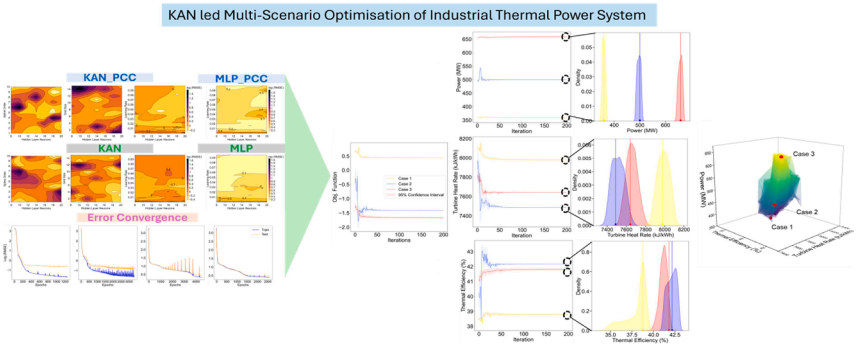
Talha Ansar [a], Waqar Muhammad Ashraf [b,c,d,*]

[a] Department of Mechanical Engineering, University of Engineering and Technology Lahore, New Campus, Kala Shah Kaku, 39020, Pakistan
[b] The Sargent Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, Torrington Place, London, WC1E 7JE, UK
[c] Collective Intelligence & Design Group, University of Cambridge, Cambridge, CB3 0HE, UK
[d] The Alan Turing Institute, British Library, 96 Euston Road, London, NW1 2DB, UK

## HIGHLIGHTS

- We present KAN_PCC algorithm – variant of KAN.
- KAN appears to be insensitive to injecting prior data information for energy systems modelling.
- KAN appears to model considered energy-systems better than MLP.
- KAN based feature importance analysis and multi-objective optimisation of thermal power system are performed.

## GRAPHICAL ABSTRACT




## ARTICLE INFO

*Keywords:*
KAN
MLP
Energy systems
Thermal power plants
Energy sustainability

## ABSTRACT

Considering the improved interpretable performance of Kolmogorov–Arnold Networks (KAN) algorithm compared to multi-layer perceptron (MLP) algorithm, a fundamental research question arises on how modifying the loss function of KAN affects its modelling performance for energy systems, particularly industrial-scale thermal power plants. In this regard, first, we modify the loss function of both KAN and MLP algorithms and embed Pearson Correlation Coefficient (PCC). Second, the algorithmic configurations built on PCC, i.e., KAN_PCC and MLP_PCC as well as original architecture of KAN and MLP are deployed for modelling and optimisation analyses for two case studies of energy systems: (i) energy efficiency cooling and energy efficiency heating of buildings, and (ii) power generation operation of 660 MW capacity thermal power plant. The analysis reveals superior modelling performance of KAN and KAN_PCC algorithms than those of MLP and MLP_PCC for the two case studies. KAN models are embedded in the optimisation framework of nonlinear programming and feasible optimal solutions are estimated, maximising thermal efficiency up to 42.17 ± 0.88 % and minimising turbine heat rate to 7487 ± 129 kJ/kWh corresponding to power generation of 500 ± 14 MW for the thermal power plant. It is anticipated that the scientific, research and industrial community may benefit from the fundamental

* Corresponding author at: The Sargent Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, Torrington Place, London, WC1E 7JE, UK.
*E-mail address:* Waqar.ashraf.21@ucl.ac.uk (W.M. Ashraf).

https://doi.org/10.1016/j.egyai.2025.100473

Available online 16 January 2025

insights presented in this paper for the ML algorithm selection and carrying out model-based optimisation analysis for the performance enhancement of energy systems.

**Nomenclature**

| | |
|---|---|
| CFR | Coal Flow Rate (t/h) |
| COBYLA | Constrained Optimization By Linear Approximations |
| CV | Condenser Vacuum (kPa) |
| ENC | Energy Efficiency Cooling |
| ENH | Energy Efficiency Heating |
| FWT | Feed Water Temperature ( °C) |
| KAN | Kolmogorov-Arnold Networks |
| KAN_PCC | Kolmogorov-Arnold Networks with Pearson Correlation Coefficient |
| MAE | Mean Absolute Error |
| ML | Machine Learning |
| MLP | Multi-Layer Perceptron |
| MLP_PCC | Multi-Layer Perceptron with Pearson Correlation Coefficient |
| MSF | Main Steam Flow rate(t/h) |
| MSP | Main Steam Pressure (MPa) |
| MST | Main Steam Temperature ( °C) |
| MW | Mega Watts |
| PCC | Pearson Correlation Coefficient |
| $R^2$ | coefficient of determination |
| RHT | ReHeat steam Temperature ( °C) |
| RMSE | Root Mean Squared Error |
| SHAP | SHapley Additive exPlanations |
| TAFR | Total Air Flow Rate (t/h) |

## Introduction

With the recent advancement in the information communication technology, internet of things and sensor technology, a large volume of data is generated across the application domains, i.e., banking [1], finance [2], industrial complexes [3,4] and social media applications [5]. The data captures the temporal-spatial dynamics bringing new insights and action plans to enhance the operation excellence and boosting efficiency for the considered application [6–9]. The ever-increasing computational power, hardware sophistication, increased literacy on the use of the data and efficient data-driven modelling algorithms, we see exponential used cases of data-driven studies promising the performance enhancement and improved understanding on how to carry out design, operation and control of considered application, say energy systems [10–16]. The increased penetration of data-driven analysis pipelines across the application domains, we are living in the fourth learning paradigm commonly referred as “data-driven sciences” where data informs the user for decision making as opposed to the computational sciences and theoretical sciences where rigorous physical laws are in play to explain the phenomenon [17].

The historic success of data-driven sciences is founded on the multi-layer feed forward perceptron based artificial neural networks which are universal function approximator and are supported by universal function approximation theorem [18,19]. Multi-layer perceptron (MLP) requires reasonable memory and compute to approximate an ill-defined function [20,21]. While, MLP models are efficient in constructing the functional mapping and digging the nonlinearity in the hyper-dimensional problem space, MLP suffers from inherent problems including sticking into the local minima, overfitting the data and black-box nature for interpretability of model-based predictions [22]. The black-box nature of MLP initially was ignored on the face of excellent modelling performance of the algorithm. However, with the increased stress on model’s transparency [23] and tracing the causal mechanism for the predictive tasks [24,25], intercepting the interpretability of MLP is a hot research question in the scientific community.

Inspired by the gap that bars the adoption of MLP for applications where model transparency and interpretation is important, recently Kolmogorov–Arnold Networks (KAN) algorithm is presented [26]. KAN is founded on the Kolmogorov–Arnold theorem that states that any multi-variate function can be approximated by sum of single-variable functions [27,28]. The theorem seems promising to interpret the nonlinear function spaces of neural networks since multivariate function space can be broken down into simple building blocks and provides improves interpretability [29]. KAN provides similar or better modelling accuracy and higher interpretability performance than those of MLP as described in original paper [26]. However, mixed results are reported in literature comparing the modelling performance of KAN and MLP for a range of real-life applications including computer vision [30], natural language processing [30], solving differential equations and operator networks [31], time-series forecasting [32], computational fluid dynamics [33] and wind power prediction [34].

KAN has been initially applied on problems having fewer input dimensions and the simulated datasets. However, it remains an open research gap as well as a promising research question to investigate how KAN’s modelling performance is comparable with those of MLP on the energy systems problems having reasonably high input dimensions and with noisy data that is a general case for the sensor based data collection [35]. Moreover, embedding the data information in the loss function of KAN, as carried out in [17], can be investigated to evaluate how data information embedding through statistical metrics affects the modelling performance of KAN since it is found useful to improve the modelling performance of neural network. While KAN has been utilised for various modelling tasks, optimisation analysis incorporating KAN model into nonlinear programming framework is not reported so far for thermal power systems. This study, for the first time, explores the effectiveness of KAN based optimisation analysis for operation optimisation of the industrial-scale thermal power system under three power generation scenarios which research and industrial community may benefit from to synergize the efforts for performance enhancement of energy systems to achieve net-zero goal.

Motivated by the fundamental as well as potential application-oriented challenges and opportunities that KAN offers, this research paper compares the modelling performance of KAN and MLP for energy systems applications taken from literature and industry. We incorporate Pearson correlation coefficient (PCC), with the rationale to infuse data information in the loss function of KAN and MLP for parameter update, and have labelled the modified algorithms as KAN_PCC and MLP_PCC which is amongst the fundamental novel aspects of this research. Then, we systematically compare the modelling performance of KAN, MLP, KAN_PCC and MLP_PCC for two case studies of energy systems: (i) energy efficiency cooling (ENC) and energy efficiency heating (ENH) of buildings [36], and (ii) performance enhancement of 660 MW capacity coal power plant. A well-trained model is deployed for SHAP based feature importance analysis [37]. Moreover, model-based optimisation analysis is also be carried out to optimize the performance parameters (power, turbine heat rate and thermal efficiency) of the power plant under different power generation modes of the power plant as framed in

[38,39]. This research investigates the comparison of KAN and MLP as well as KAN_PCC and MLP_PCC for optimisation of coal power plant's operation for the first time as per the best knowledge of authors. The modelling comparison of the two algorithms provide fundamental insights about the suitability of these algorithms for the analysis of energy systems that may guide the practitioners and researchers to well design the ML based methodologies for the analysis of energy system.

This research paper is structured as follows: First we describe the algorithmic working of KAN and MLP that also includes the details of different hyperparameters associated with each algorithm. Next, we modify the loss function of both KAN and MLP with PCC term that contributes towards parameters update of the two algorithms. The modified version of KAN and MLP is labelled as KAN_PCC and MLP_PCC. The systematic comparison of KAN, KAN_PCC, MLP and MLP_PCC is made for modelling the output variables of the two case studies of energy systems. Later, a well-predictive algorithm is coupled with SHAP technique to establish the feature importance towards the performance parameters of thermal power plants. Lastly, KAN is integrated into optimisation framework to estimate the optimal operating conditions for power generation operation of thermal power plant. The analysis carried out in this paper is implemented in Python using different packages suitable for the tasks.

## Method

### *Kolmogorov-Arnold network*

The Kolmogorov-Arnold representation theorem states that multivariate functions can be decomposed into a sum of univariate functions. Specifically, for a smooth function $f: [0, 1]^n \rightarrow \mathbb{R}$, it can be represented as:

$$f(x) = f(x_1, \ldots. x_n) = \sum_{q=1}^{n+1} \Phi_q \left( \sum_{p=1}^{n} \Phi_{q,p}(x_p) \right) \tag{1}$$

where $\left(\phi_{q,p} : [0,1] \rightarrow R\right)$ and $\left(\Phi_q : R \rightarrow R\right)$. This representation implies that the only true multivariate function is addition, as every other function can be expressed using univariate functions combined through summation. A KAN layer with $(n_{in})$-dimensional inputs and $(n_{out})$-dimensional outputs can be characterized as a matrix of 1D functions:

$$\left[\Phi = \phi_{q,p}, \quad p = 1, 2, \ldots, n_{in}, \quad q = 1, 2, \ldots, n_{out}\right] \tag{2}$$

where the functions $\left(\phi_{q,p}\right)$ contain trainable parameters. In the context of the Kolmogorov-Arnold theorem, the inner functions establish a KAN layer with $(n_{in} = n)$ and $(n_{out} = 2n + 1)$, while the outer functions create a KAN layer with $(n_{in} = 2n + 1)$ and $(n_{out} = 1)$. A Kolmogorov-Arnold Network (KAN) explicitly parameterizes the representation, where each 1D function is represented as a B-spline curve with learnable coefficients. The architecture of a KAN can be denoted by an integer array:

$$\left[n_0, n_1, \ldots, n_q\right],$$

where $(n_i)$ represents the number of nodes in the $(i)$-th layer. The KAN output, given an input vector $(x_0 \in R^{n_0})$, is defined as:

$$KAN(x) = \left(\Phi_{q-1} \circ \Phi_{q-2} \circ \ldots \circ \Phi_1 \circ \Phi_0\right) x \tag{3}$$

Here, $x$ represent the set of input variables with $N$ observations. KAN consists of multiple computational layers, where each layer comprises a matrix of trainable univariate functions $\left(\phi_{q,p}\right)$. The primary computation in the KAN is expressed in Eq. (1), where $(p_k)$ denotes the summation for the $(k)$-th layer, $(n_{in})$ is the number of input dimensions, and $(n_{out})$ is the number of output dimensions. This summation is then transformed using an activation function $(f_k)$ to yield the layer output$(y_k)$. The model-based response $(Z)$ corresponding to the input condition $(X)$ is computed as

$$Z = \Phi_q\left(\Phi_{q-1}\left(\ldots\left(\Phi_1\left(y_0\right)\right)\ldots\right)\right)X \tag{4}$$

where $(q)$ denotes the total number of layers. The forward pass processes the information through these layers, while during training, the error between $Z$ and the true response $(D)$ is backpropagated to update the KAN parameters, enhancing the model's performance. More details about the working of KAN can be found in [26].

The algorithm of KAN is embedded with a large number of hyperparameters to tune the predictive performance. The hyperparameters of KAN include the grid size, which is selected from a quantized uniform distribution ranging [1,15], defining the resolution of the grid used in the model. The spline order is drawn from a quantized uniform distribution between [1,10], specifying the degree of the spline used for interpolation. The scale of noise is sampled from a log-uniform distribution within the range $\left[10^{-5},\ 10^{-1}\right]$, controlling the magnitude of noise added to the model. The scale base is also drawn from a log-uniform distribution ranging $\left[10^{-5},\ 10^{1}\right]$, determining the base scaling factor for model inputs. The scale of the spline is sampled from a log-uniform distribution in the range $\left[10^{-5},\ 10^{1}\right]$, influencing the scaling of the spline used in the model. The learning rate is sampled from a log-uniform distribution in the range $\left[10^{-5},\ 10^{-1}\right]$, controlling the step size at each iteration during the optimization process. The $L_1$ regularization term $(\lambda_1)$ is drawn from a log-uniform distribution spanning $\left[10^{-7},\ 10^{-1}\right]$, adding a penalty to the loss function based on the absolute values of the coefficients to promote sparsity. The weight decay term $(\lambda_2)$ is also sampled from a log-uniform distribution within the range $\left[10^{-7},\ 10^{-1}\right]$ contributing towards the parameters update. The parameter grid epsilon is sampled from a log-uniform distribution within the range $\left[10^{-7},\ 10^{-1}\right]$, determining the threshold for grid refinement during optimization. The number of neurons in the hidden layer is selected from a quantized uniform distribution with values ranging between 8 and 20 in steps of 2, defining the architecture complexity of the KAN. Finally, the activation scale, which adjusts the scaling of the activation functions to control the intensity of non-linear transformations in the model, is sampled from a uniform distribution ranging [1,10]. The hyperparameters are tuned with HypertOpt package in Python which incorporates Tree-structured Parzen Estimator for sequential model based optimisation of hyperparameters [40]. Moreover, HyperOpt can handle asynchronous distribution space for optimisation of hyperparameters [41].

### *Multi-layer perceptron*

A feed forward multi-layer perceptron model is inspired from the working of neurocomputations, taking place in the human brain. The algorithm propagates the input data into the computing layer where it is transformed and transmitted to the output layer for further processing until a response is generated. Mathematically, the working of MLP is represented as:

$$p_1 = \sum W_1 \odot X^T + b_1 \tag{5}$$

$$y_1 = f_1(p_1) \tag{6}$$

here, $X$ is the set of input variables having $N$ observations; $W_1$ is the weight connections from the input to hidden layer of MLP and $b_1$ is bias introduced on the nodes of hidden layer of MLP. A summation is computed $(p_1)$ on each hidden layer neuron of MLP using data processing $(\sum W_1 \odot X^T + b_1)$ and is transformed by activation function $f_1$. Activation function transforms $p_1$ into nonlinear function space and computes a signal strength of $y_1$ that is further processed at the output layer of MLP using Eqs. (7) and (8), and finally a model-based response $(Z)$ is simulated corresponding to input conditions of $X$

$$p_2 = \sum W_2 \odot y_1 + b_2 \tag{7}$$

$$Z = f_2(p_2) \tag{8}$$

here, $W_2$ is the weight connection values from hidden to output layer of MLP; $b_2$ is the bias embedded at the output layer neuron and $f_2$ is the activation function that further processes $p_2$. The information processing from Eqs. (5) to (8) represents the feed forward information pass for simulating $Z$. In the back propagation, the error is computed between the $D$ and $Z$ and is propagated backwards to update the MLP parameters ($\Theta$) to achieve good modelling performance. Reader may find more details about the working of MLP in [42].

The hyperparameters of the MLP are also tuned by HyperOpt package, and include the learning rate, which is sampled from a log-uniform distribution within the range [$10^{-5}$, $10^{-1}$], and the $L_1$ regularization term ($\lambda_1$) and weight decay term ($\lambda_2$), both drawn from log-uniform distributions spanning [$10^{-7}$, $10^{-1}$]. The number of neurons in the hidden layer is selected from a quantized uniform distribution with values ranging between 8 and 20 in steps of 2. The activation scale, which adjusts the scaling of the activation functions to control the nonlinear intensity, is sampled from a uniform distribution ranging in the range [1,10].

*Loss function*

The loss function ($\mathscr{L}$) computes the error between $D$ and $Z$ and is customized to construct the variants of KAN and MLP for this research. Traditionally, $\mathscr{L}$ is defined on mean square error term and $L_1$ regularization that we have also implemented for parameters update of KAN and MLP. However, $\mathscr{L}$ is also customized with Pearson correlation coefficient (PCC, $r$) that is deployed for training of KAN_PCC and MLP_PCC and is given as:

$$\mathscr{L} = \frac{\sum (D - Z)^2}{N} + \lambda_1 \Theta + \rho \frac{\sum_{i=1}^{m} \left(r_{X_i|D} - r_{X_i|Z}\right)^2}{N} \tag{9}$$

here, $\Theta$ is the model parameters that are minimized for feature selection of ML models. The PCC is introduced as to minimize the mean square deviation between $r_{X_i|D}$ and $r_{X_i|Z}$ and contributes towards the parameters update of KAN and MLP. It is important to mention here the rationale of brining PCC into $\mathscr{L}$ is that PCC computes the linear association between the two pair of variables; this is the information that is available from the data exploratory analysis and depicts the characteristics of the system as depicted through the statistical term. Thus, similar to Physics Informed Neural Network, where physical model is also introduced in the $\mathscr{L}$ to update the model parameters [43], data metrics computed on the quality data represents the data-driven insights available that are embedded in the algorithm design for the parameters tuning. Thus, it is anticipated that data metrics including PCC may infuse interpretability in the typical black box models including MLP and this approach has shown positive outcome in the interpretation analysis as reported in [17].

The customized $\mathscr{L}$ is embedded in KAN_PCC and MLP_PCC architecture. The penalty factor ($\rho$) for the PCC loss term, used in $\mathscr{L}$, is sampled from a log-uniform distribution in the range $[10^{-7}, 10^{1}]$, adjusting the impact of the penalty term in the PCC loss function for KAN_PCC and MLP_PCC models. The model parameters are tuned by Adaptive Moment Estimation (ADAM) solver. The solver leverages the characteristics of RMSprop and stochastic gradient descent for the error convergence, computationally efficient, has little memory requirement and invariant to diagonal rescaling of gradients [44].

*Model evaluation*

The models are evaluated on three statistical features, relevant to the regression problems, namely coefficient of determination ($R^2$), root-mean-squared-error (RMSE) and mean absolute error (MAE). The mathematical expressions of $R^2$ and RMSE are given as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \widehat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \overline{y}_i)^2} \tag{10}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\widehat{y}_i - y_i)^2} \tag{11}$$

$$MAE = \frac{\sum_{i=1}^{N} |(\widehat{y}_i - y_i)|}{N} \tag{12}$$

here, $y_i$ and $\widehat{y}_i$ are the true and model-predicted values; $\overline{y}_i$ is the mean of $y_i$; $i$ = 1,2,3,…, $N$ equal to number of observations. $R^2$ measures the proportion of variance in $y_i$ captured by $\widehat{y}_i$; thus, model-based well-predicted values have high value of $R^2$ reaching to 1. Whereas, a poor-predictive model depicts lower $R^2$ value reaching to zero. On the other hand, RMSE and MAE measure the mean error in the model predicted responses; higher RMSE and MAE are the indicator of low-predictive accuracy of the trained model; thus, the errors are expected to be as low as possible to ensure good fit on data.

Inductive Conformal Prediction (ICP) technique is applied to compute the prediction intervals on the test dataset. The ICP provides validated prediction intervals and effectively quantifies the uncertainty bounded with the trained ML model, providing a competitive edge over other uncertainty quantification techniques including Bayesian and ensemble methods [45]. Following the training of the ML models, ICP method is utilised to construct the prediction intervals on 95 % confidence level by ICP technique for uncertainty estimation on the point-prediction made by the ML model. The reader may consult with for understating the working of ICP technique [46].

*SHAP based feature importance analysis*

SHAP (SHapley Additive exPlanations) is a versatile and robust technique to investigate the interpretation performance of the ML models. SHAP is model-agnostic technique based on game theory [47], where SHAP values are calculated by creating instances that simulate the impact of input variables on the target variable. Many techniques for feature importance analysis exist such Monte Carlo method [48], one-factor-at-a-time analysis [49], and the partial derivative method [50]. Among the mentioned techniques, SHAP investigates the interpretation performance from the dimension of efficiency, symmetry, dummy, and additivity [51]. SHAP technique offers two important features, local accuracy and consistency. Moreover, SHAP provides insights into model sensitivity to changes in input variables, at a local level (specific data points) and a global level (the entire dataset) determining the importance and ranking of input variables. Feature importance analysis is important in the ML based analysis framework since ML models are treated as black-box functions requiring the need to interpret the model prediction based on significant features [52]. Furthermore, KAN model will be also integrated in the SHAP framework to establish the feature importance for the input variables of the case studies.

*Formulating optimisation problem and estimating solution uncertainty*

The nonlinear variables interactions and nonlinear function profiles of the performance variables for energy systems like thermal efficiency and turbine heat rate direct to apply nonlinear programming technique to formulate the optimisation problem [53]. Nonlinear programming framework can effectively estimate the solution for the nonlinear objective or nonlinear constraints introduced in the optimisation problem. A general mathematical framework for multi-objective optimisation problem is given as:

$$\begin{aligned}&\text{Multi}-\text{objective function}: \min_x f(x) = (f_1(x) + f_2(x) + \ldots + f_c(x))\\ &\text{Subjectto}:\\ &h(x) = 0\\ &x = x_1,\ x_2, \ldots, x_m\\ &x\ \epsilon\ X\ \subseteq\ R^n\\ &x^L\ \leq x \leq x^U\end{aligned} \tag{13}$$

here, $f_1(x) + f_2(x) + \ldots + f_c(x)$ is $c$ number of objective functions incorporated to define the multi-objective function $f(x)$. $+$ sign means that the objective function is to be minimized while function to be maximized is written with negative sign in $f(x)$. $h(x)$ is the linear constraint representing ML models which are MLP and KAN in this research. $x$ is the set of $m$ input variables the ML models are trained on. Moreover, $x$ is continuous and real in nature and is bounded on lower ($x^L$) and upper limit ($x^L$) that serves as design space for solution estimation. More details about nonlinear programming technique can be found in [54].

We have utilized COBYLA solver for estimating the local optimal solution for $f(x)$ [55]. Considering the local solutions obtained from the solver, we have tried multiple initial guesses within the design space of $x$ to estimate feasible solution ($x^*$) for $f(x)$ and have estimated the prediction intervals with inductive conformal prediction technique [56].

The uncertainty in $x^*$ is quantified through Monte Carlo simulations. The solution space is perturbed through uniformly distributed noise generated on one percent value of the input variables ($\delta_k,\ k = 1,2,3,\ldots, H$). The simulated experiments, constructed on $x^*$ and noise observations, are simulated by the trained ML models, and the variance in the estimated solution ($V(x^*)$) is computed to account for the robustness of the solution on the face of random variation in the input variables. The mathematical expressions computing $V(x^*)$ in the solution are given as:

$$V(x^*) = \frac{\| F(x^*) - f(x^*)\|}{\| f(x^*)\|} \tag{14}$$

where,

$$F(x^*) = \frac{\sum_{k=1}^{H} f(x^* + \delta_k)}{H} \tag{15}$$

The mean response ($F(x^*)$) of the function space is computed through Monte Carlo experiments ($x^* + \delta_k$) that contributes to calculate variance in the vicinity of local function profile, measured against the perturbed inputs. The constructed Monte Carlo experiments propagate the uncertainty in the input variables to the predicted responses from the ML model-based function. Lower is the value of variance against the perturbed inputs, the estimated solution is robust to the face of uncertain inputs.

## Results & discussion

### *Case-1: modelling energy efficiency cooling and energy efficiency heating of buildings*

The dataset for energy efficiency cooling (ENC) and energy efficiency heating (ENH) is compiled for the energy efficiency performance of the buildings. The dataset contains eight building performance related features namely orientation, roof area, wall area, surface area, glazing area distribution, relative compactness, glazing area and overall hight. The dataset has compiled 768 observations corresponding to the input variables considering the geometrical features of the buildings. 80 % of the data is partitioned for the training of the model while remaining 20 % is utilized to test the modelling performance of the models to predict ENC and ENH.

The hyperparameters are tuned on the wide operating ranges, considering different distribution profile suitable for parametric variation for efficient parameter estimation for KAN, MLP, KAN_PCC and MLP_PCC. Log (RMSE) is computed on the test dataset against each combinatorial values of hyperparameters and the resulting error profile is presented on Fig. 1 for ENC. Contour plots are constructed on hidden layer neurons against spline order, grid size and learning rate depicting log(RMSE) error profile for KAN_PCC as shown on Fig. 1(a). Multiple dips and numerous local minima in the error profile of KAN_PCC is observed, depicting the significance of the hyperparameters towards the curve-fitting across the data. The error profile of KAN_PCC appears to be sensitive to combinatorial values of hyperparameter, while non-convex and highly nonlinear function space is observable for KAN_PCC model. The constructed contour plots confirm the high sensitivity of spline order, grid size and learning rate towards the test error. Similar is the case for error distribution against the parametric ranges of KAN based hyperparameters as shown in Fig. 1(b), signifying the sensitivity of the algorithm in locating the error minimum in the deep error valley characterised by numerous error minima. Hidden layer neurons and learning rate turn out to be significant hyperparameters for MLP_PCC and MLP respectively as shown in Fig. 1(c-d). Hidden layer neurons introduce the complexity to approximate the function on data-centric learning while learning rate adjusts the step size to reach the error minimum during the learning phase of the model. The hyperparameter values of KAN_PCC, KAN, MLP_PCC and MLP are selected corresponding to the lowest RMSE computed on the test dataset, and are embedded in the model's architecture. The models are trained on the optimal values of hyperparameters to update the model-based parameters by optimizer to finally develop a well-predictive model approximating the underlying function in the dataset.

Fig. 1(e) shows the error convergence against the iterative training of the models computed on the training and testing dataset. The log (RMSE) error is reduced nearly smoothly on training and testing dataset for KAN_PCC as shown on Fig. 1(e)(i) where training error continues to improve while testing error is improved marginally over the iterative training. This shows that sufficient iteration budget has been allocated for training the KAN_PCC algorithm and bias-variance trade-off is also reasonably small, displaying good training characteristics of the model. The loss function-based error is gradually reduced on training and testing dataset for MLP_PCC, KAN and MLP and nearly stabilized on training and testing dataset as shown on Fig. 1(e)(ii-iv), respectively. The gradual reduction in loss function error shows the error convergence direction, the algorithms are following during the model training phase. The error convergence for MLP_PCC has fluctuating trend during the model training phase. However, overall reduction in error is observed during the iterative training of the model. KAN model also depicts the sharp peaks in the error profile, observed for training and testing datasets, which is explained by the ineffective parametric update deteriorating the model performance. It is also important to note here that testing error is gradually and marginally improved over training episodes and the error divergence is not observed, meaning the risk of overfitting is relatively small in the trained models. Similarly, the hyperparameters are also tuned for training the models for ENH. The contour plots against the hyperparameters of the models and the error convergence during the model training for ENH are shown in Fig. A1 in supplementary file.

The well-trained models corresponding to optimal values of hyperparameters and rigorous training during the model development phase are retained for ENC and ENH. The parity plots mapping the true and model predicted values of ENC and ENH are shown on Fig. 2. The modelling performance is computed on the training and testing datasets, and the distribution profiles of the datasets are also shown along the edges. Referring to ENC on Fig. 2(a), KAN_PCC model demonstrates the same $R^2$ values, i.e., 0.99 on training and testing datasets. KAN_PCC is found to have RMSE and MAE errors of 0.18 % and 0.14 % respectively on training while 0.52 % and 0.39 %, respectively on testing datasets respectively. Whereas, KAN has RMSE_train and MAE_train errors of 0.19 % and 0.15 %, respectively. Moreover, KAN-based RMSE_test = 0.42 % and MAE_test = 0.31 % are lower than those of KAN_PCC on test dataset for ENC. The comparative modelling performance of MLP_PCC

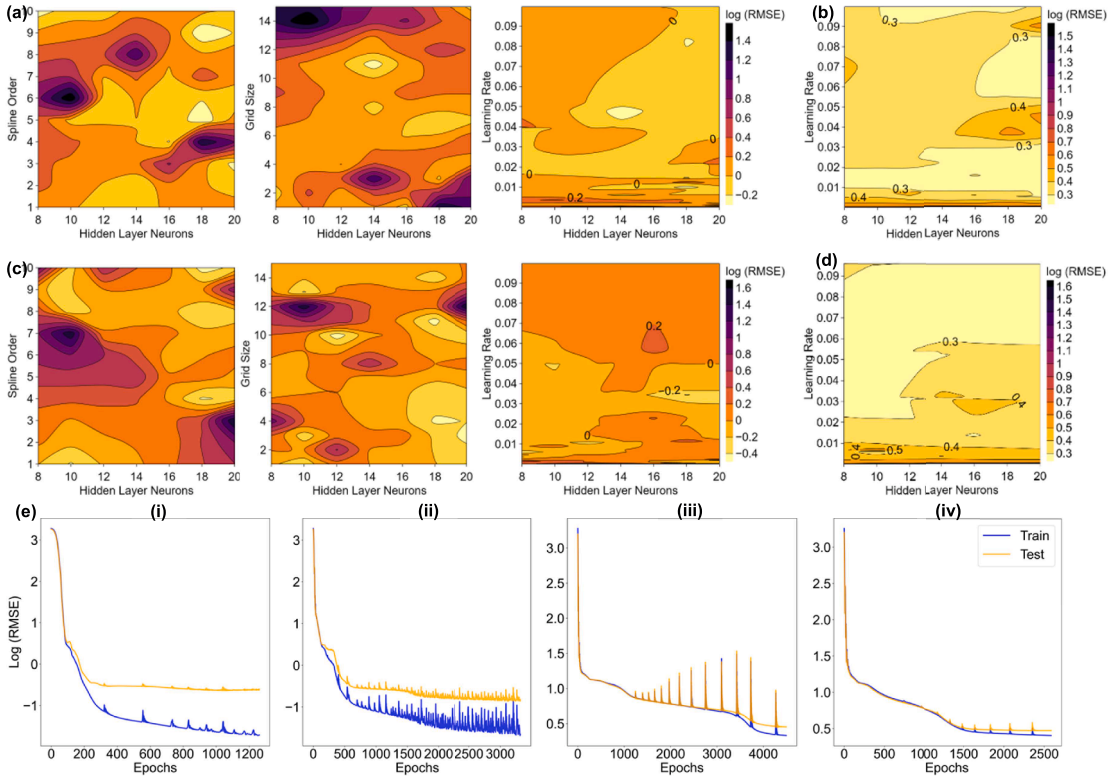



**Fig. 1.** Training the KAN and MLP models with and without PCC in the loss function for ENC. (a) Mapping the loss function error (log (RMSE) against the hyperparameter for (a) KAN_PCC, (b) MLP_PCC, (c) KAN and (d) MLP. (e) The error convergence profiles on training and testing dataset are shown for (i) KAN_PCC, (ii) MLP_PCC, (iii) KAN and (iv) MLP, respectively.

and MLP reveals that the two models have same $R^2$ value of 0.97 both on training and testing datasets. MLP_PCC has RMSE_train = 1.42 % and MAE_train = 0.96 % which are lower than those of MLP (RMSE_train = 1.50 %, MAE_train = 1.00 %). While, RMSE and MAE computed on test dataset for MLP_PCC, i.e., 1.56 % and 1.04 % are slightly better than those of MLP to predict ENC.

Similar observations are noted for comparing the modelling performance of KAN and KAN_PCC for ENH as shown on Fig. 2(b) where KAN seems to outperform KAN_PCC, both on training and testing datasets. Whereas, MLP_PCC and MLP has same performance metrics on the training and testing datasets to predict ENH. The modelling comparison of KAN with respect to incorporation of PCC in the loss function reveals that PCC seems not to improve the modelling performance of KAN for ENC and ENH models. This behaviour of KAN algorithm can be explained by the nature of model architecture to depict interpretability in the responses. However, the PCC information seems to provide marginal improvement in the modelling performance for ENC and ENH as demonstrated in the performance metrics.

*Case-2: modelling and optimisation of performance parameters of thermal power plant*

### i- Modelling Power, Turbine Heat Rate and Thermal Efficiency of thermal power plant by KAN_PCC, KAN, MLP_PCC and MLP

The comparison of KAN and MLP with and without the incorporation of PCC is made for industrial application on the power generation operation of a 660 MW capacity thermal power plant. The case study is chosen to investigate the comparative performance analysis for the real-data of the industrial systems and the suitability of the algorithms for the modelling tasks.

Three plant-level performance parameters including Power, thermal efficiency and turbine heat rate are modelled on the input variables by KAN and MLP. The dataset is taken from [10] that has 1278 observations associated with the input-output variables. The input variables include coal flow rate (CFR – t/h) and total air flow rate (TAFR – t/h), main steam pressure (MSP – MPa), main steam temperature (MST – °C), main steam flow rate (MSF – t/h), feed water temperature at the boiler's inlet (FWT – °C), reheat steam temperature (RHT – °C) and condenser vacuum (CV – kPa). The hyperparameters are tuned on the design space and parametric distribution, and are evaluated on the testing dataset (0.2 split ratio of 1278 observations).

Fig. 3 depicts the hyperparameters tuning and error convergence profiles during the training of ML models for Power. Referring to Fig. 3 (a), numerous local valleys are observed on the error space constructed on the hyperparameters including hidden layer neurons with spline order and grid size for KAN_PCC for Power. This shows the highly parametric sensitivity of the KAN_PCC in minimizing the error. Lower parametric values of learning rate seem favourable for KAN_PCC that tends to minimize the error. Similar behaviour of error space is observed for learning rate for MLP_PCC model where learning rate up to 0.01 seems to impact significantly to error computed during combinatorial values of the hyperparameters, as shown in Fig. 3(b). We notice similar impact of the hyperparameters on the error space constructed for KAN and MLP, as observed for KAN_PCC and MLP_PCC, and have presented the influence of the hyperparameters on the two models on Fig. 3(c,d), respectively. The optimal values of the hyperparameters are selected on the lowest test RMSE computed during the hyperparameters tuning, and are embedded for the training of KAN_PCC, MLP_PCC, KAN and MLP models.

The loss function-based error convergence profile on the training and testing datasets for the ML models is depicted on Fig. 3(e). Smooth error convergence trend is observed during the iterative training of KAN_PCC, MLP_PCC, and KAN as represented on Fig. 3(i-iii), respectively indicating the sequential improvement in the learning process on the parametric updates. The error computed on loss function is decreasing

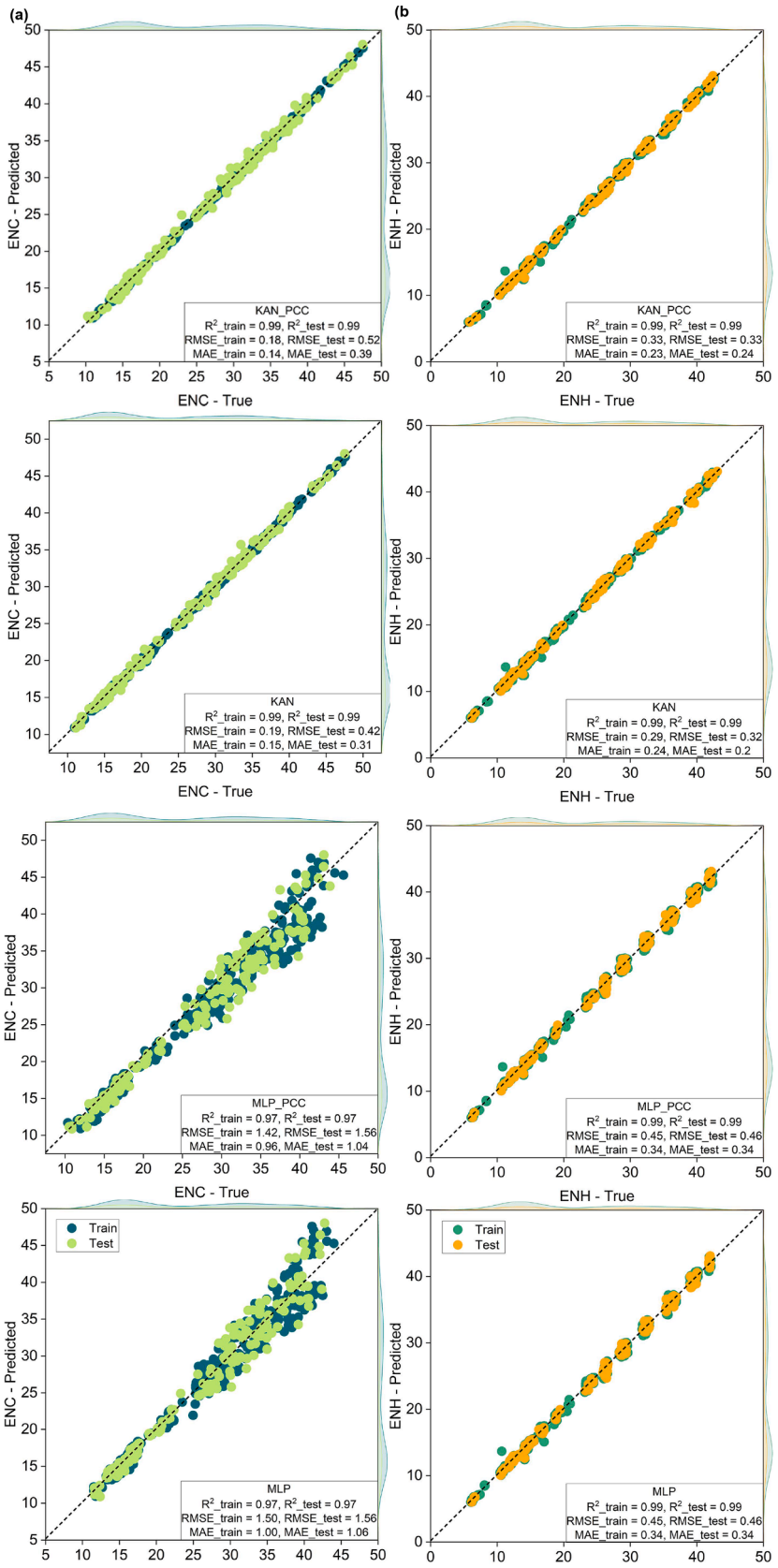



**Fig. 2.** Modelling performance of the KAN_PCC, KAN, MLP_PCC and MLP models on the training and testing datasets for (a) ENC and (b) ENH, respectively. The true and model-based data distribution profiles are also shown along the edges of the figures.

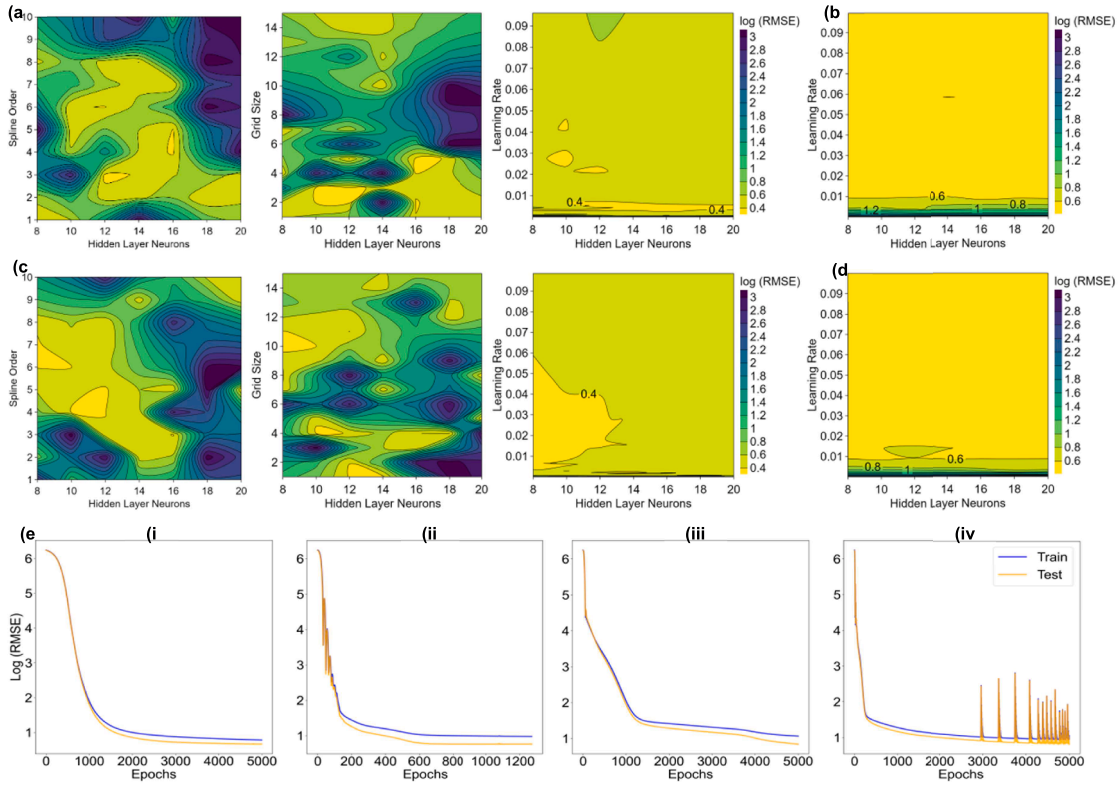



**Fig. 3.** Training the KAN and MLP models with and without PCC in the loss function for Power. (a) Mapping the loss function error (log (RMSE) against the hyperparameter for (a) KAN_PCC, (b) MLP_PCC, (c) KAN and (d) MLP. (e) The error convergence profiles on training and testing dataset are shown for (i) KAN_PCC, (ii) MLP_PCC, (iii) KAN and (iv) MLP, respectively.

for MLP over the training iterations; however, error spikes and fluctuations are also observed near the completion of iteration budget for the MLP model training as shown on Fig. 3(e)(iv). This shows the unstable learning process for MLP near the utilization of the iterative budget and ineffective update in the model-based parameters that adversely impact the performance measuring metrics. The hyperparameters are also comprehensively explored and rigorously tuned for the training of ML models for turbine heat rate and thermal efficiency taken from the power plant's operation. The plots for hyperparameters tuning and error convergence for turbine heat rate and thermal efficiency are depicted on Figs. A2 and A3, respectively in the supplementary file.

Fig. 4 compares the modelling performance of ML models on the training and testing datasets for the three performance parameters of the thermal power plant. The distribution profiles for true and model-predicted responses are also plotted along the edges of the figures to compare the distribution similarity for the datasets. Referring to Fig. 4 (a), $R^2$ of 0.99 is computed for training and testing datasets for KAN_PCC and KAN models for Power. KAN based RMSE and MAE on training dataset is 2.05 MW and 1.40 MW which are lower than those of KAN_PCC (RMSE_train =2.19 MW, MAE_train = 1.49 MW). However, KAN_PCC appears to have marginally improved predictive performance on test dataset with RMSE_test and MAE_test of 1.94 MW and 1.44 MW respectively which are lower than those of KAN (RMSE_test = 1.96 MW and MAE_test = 1.39 MW). Whereas, nearly similar performance metrics are computed to predict Power on testing dataset for MLP_PCC and MLP with $R^2$ of 0.99 and RMSE_test = 2.25 MW and MAE_test = 1.66 MW. Comparing the predictive performance of KAN and KAN_PCC with MLP and MLP_PCC, KAN_PCC appears to provide superior modelling performance to predict power with lower RMSE and MAE on test dataset than those of MLP.

The predictive performance of KAN, MLP and the proposed KAN_PCC and MLP_PCC models for turbine heat rate is presented on Fig. 4(b). KAN_PCC seems to have marginally improved predictive performance on training dataset ($R^2 = 0.87$, RMSE = 62 kJ/kWh, MAE = 45 kJ/kWh) than KAN ($R^2 = 0.86$, RMSE = 65 kJ/kWh, MAE = 47 kJ/kWh). However, the two models demonstrate nearly the same predictive performance on the test dataset ($R^2 = 0.86$, RMSE = 60 kJ/kWh, MAE = 46 kJ/kWh). Whereas, MLP_PCC and MLP have almost the same performance metrics computed on training ($R^2 = 0.83$, RMSE = 71 kJ/kWh, MAE = 52 kJ/kWh) and testing datasets ($R^2 = 0.86$, RMSE = 61 kJ/kWh, MAE = 48 kJ/kWh) to predict turbine heat rate. The performance comparison of KAN_PCC, KAN, MLP_PCC and MLP reveals that KAN_PCC has relatively improved modelling performance to predict turbine heat rate. Similar performance behaviour of KAN and MLP is observed to predict thermal efficiency where KAN performance on test dataset ($R^2 = 0.91$, RMSE = 0.29 %, MAE = 0.21 %) is comparatively better than those of MLP ($R^2 = 0.87$, RMSE = 0.36 %, MAE = 0.27 %). Overall, KAN seems to provide improved modelling performance to predict Power, turbine heat rate and thermal efficiency than MLP and the models are deployed for subsequent analysis as presented in the following.

### ii- Uncertainty Quantification for ML model based point-predictions

Following the training of ML models and their evaluation on performance measuring metrics, the ICP technique is applied to construct the prediction intervals around the model-based point predictions. The prediction intervals estimate the range of variation in the model-based point prediction, quantifying the uncertainty for the predictive tasks. The prediction intervals are drawn around the true test observations to validate their data-coverage accuracy when ML model simulates the unseen input conditions.

Fairly tight prediction intervals are constructed on true-test observations for power (MW), as shown on Fig. 5(a), indicating the good modelling accuracy of the trained KAN model. The prediction intervals computed for turbine heat rate (kJ/kWh) and thermal efficiency (%), as

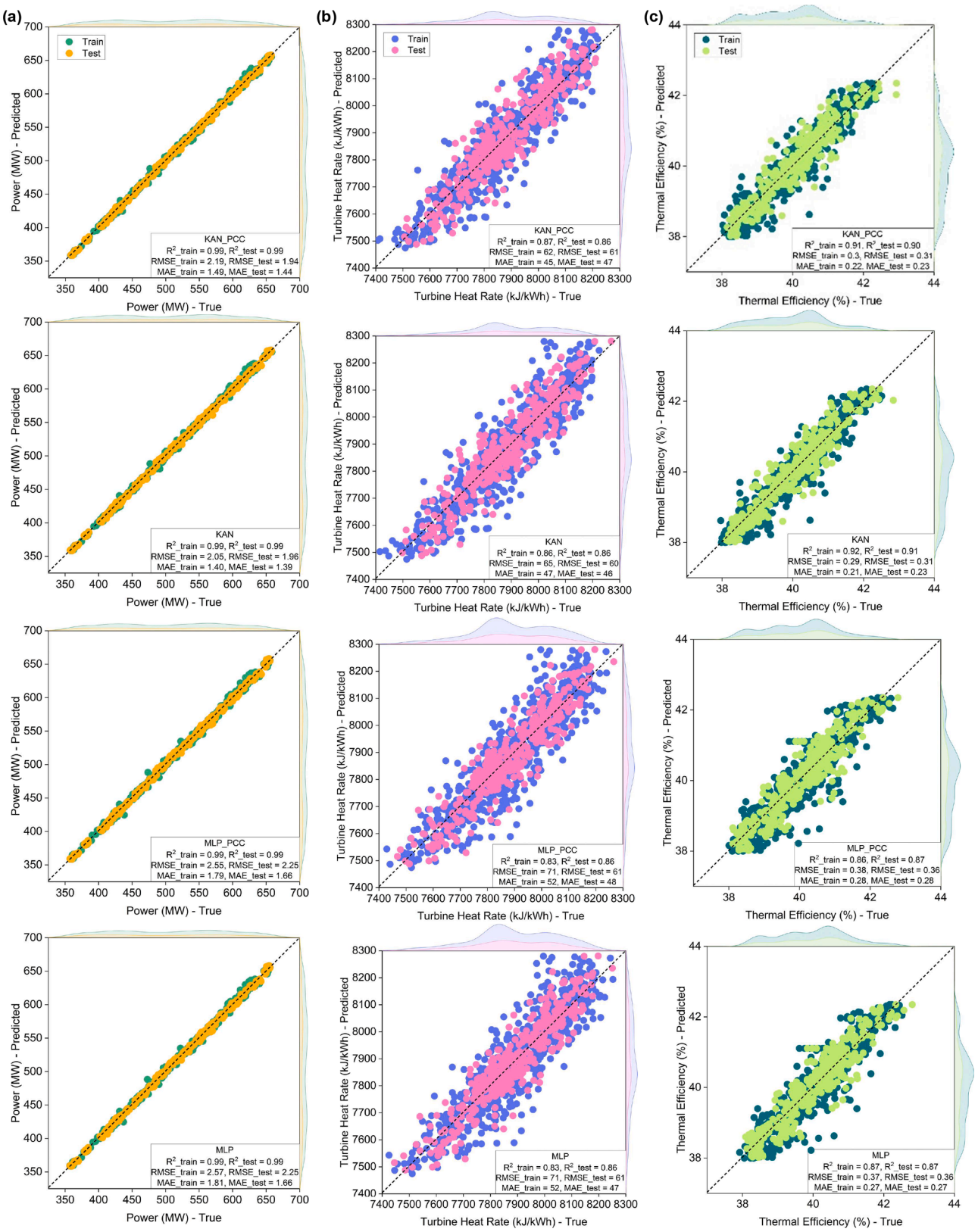



**Fig. 4.** Modelling the performance parameters of thermal power plant by KAN and MLP. The predictive performance is measured on training and testing dataset for KAN_PCC, KAN, MLP_PCC and MLP for (a) Power (MW), (b) Turbine heat rate (kJ/kWh) and (c) Thermal efficiency (%).

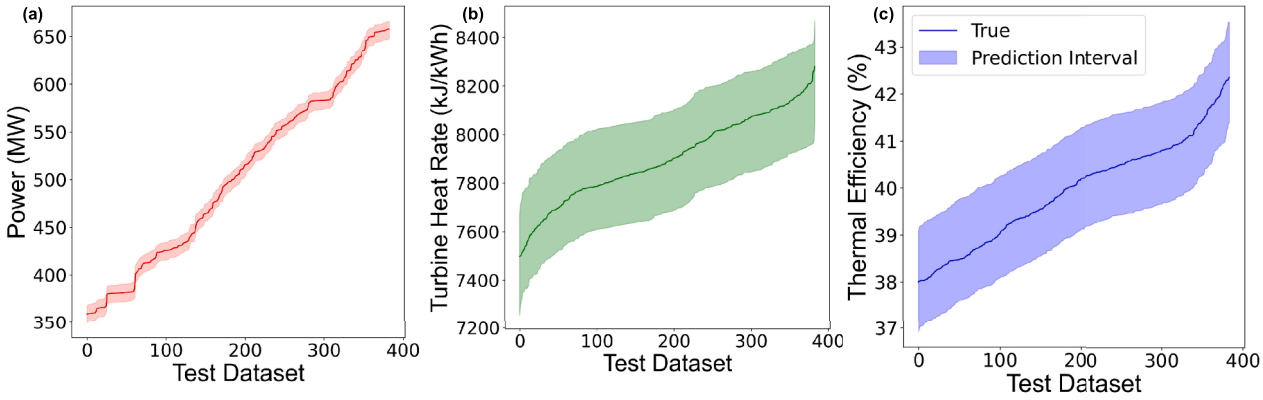



**Fig. 5.** Prediction interval estimation by inductive conformal prediction technique for (a) Power (MW), (b) Turbine Heat Rate (kJ/kWh), and (c) Thermal Efficiency (%).

shown on Fig. 5(b) and Fig. 5(c), respectively, appear to be slightly wide; yet effective uncertainty quantification is pivotal for informed decision-making when confronted with uncertain input conditions and capacity of the ML algorithms for modelling complex function spaces of the real-life applications.

### iii- Feature Importance Analysis

KAN fits the splines of different order across the data to approximate the function and it becomes imperative to investigate the feature importance analysis, impacting the predictions made by KAN. SHAP is used for feature importance analysis for the trained KAN models as it provides good estimate of the order of features, meaning interpretability performance for the regression-based ML models. The percentage contribution of features impacting power, turbine heat rate and thermal efficiency is mentioned in Fig. 6(a-c). FWT, MSFR and MSP are found to be first three significant features for predicting power having percentage contribution of 43.41 %, 28.78 % and 9.19 %, respectively. The three features, i.e., FWT, MSFR and MSP are also observed to be the same three significant features for turbine heat rate and thermal efficiency having percentage contribution of 37.71 %, 16.15 %, 14.49 % and 40.52 %, 33.24 %, 12.56 %, respectively.

FWT is measured at the entrance of boiler and is maintained on heat exchange, by steam extractions from steam turbine systems, with the feed water passed through the feed water regenerative system. The fuel consumption or energy spent in the boiler is conditioned on the thermodynamic state or conditions of FWT and the variation in FWT may result in disturbance in the boiler operation through abrupt change in the fuel injection by the installed control system and heat exchange in the heating surfaces including water walls, superheater, reheater, economizer etc. is affected for the steam generation. The significance of FWT during the plant operation is also identified by KAN based feature importance analysis that aligns with the domain-knowledge of the plant's operation. The state of power generation is effectively controlled by the steam conditions at the entrance of steam turbine systems which are measured in MSFR and MSP. MSFR is the amount of steam injected in the steam turbine system while the work potential of the steam is effectively adjusted with MSP in the boiler. The two thermodynamic parameters are synchronized with the ramp-up and ramp-down operation in the plant's operating environment and have significant impact to control turbine heat rate and thermal efficiency during the plant operation.

### iv- KAN based Optimisation Analysis of Thermal Power Plant

The KAN models are trained with good modelling accuracy for the function approximation and to predict power, turbine heat rate and thermal efficiency of thermal power plant. The function profile of thermal efficiency and turbine heat rate is continuous, nonlinear and non-convex, so nonlinear programming technique is suitable for formulating the multi-objective optimisation problem. The multi-objective optimisation problem is defined to maximize thermal efficiency and minimize turbine heat rate corresponding to three power generation scenarios. The three power generation scenarios are considered corresponding to power interval of 350 MW – 360 MW (case-1), 490 MW – 500 MW (case-2) and 650 MW – 660 MW (case-3). KAN

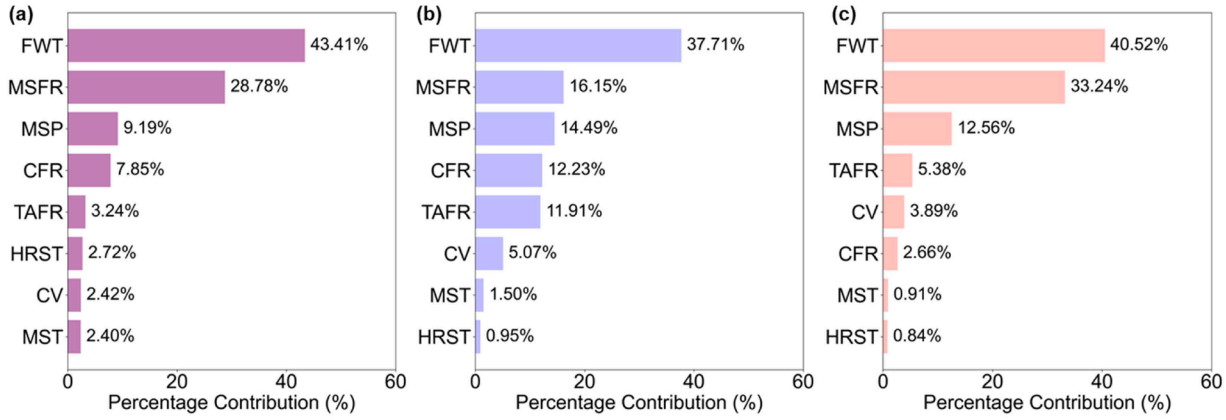



**Fig. 6.** KAN driven feature importance analysis for (a) Power, (b) Turbine heat rate and (c) Thermal efficiency. FWT, MSFR and MSP are the first three significant variables impacting the predictions made by KAN models for three performance parameters.

models are introduced as equality constraints and are also embedded in the multi-objective function. The bounds are provided on the search space of input variables corresponding to the three states of power generation operation of thermal power plant. COBYLA solver is utilized for solution estimation corresponding to the different initial guesses. A large number of initial guesses are tried to determine feasible solution for the performance parameters. Moreover, solver-driven uncertainty and uncertainty in the determined solutions are also investigated.

Fig. 7 displays the iterative progress of the solver to estimate the optimal solution for the multi-objective optimisation problem. The three cases of power generation are introduced as constraint in the optimisation problem for the optimal solution estimation to optimise the performance of the power plant. Fig. 7(a) shows the improvement in the objective function during the iterative improvement in the solution estimation for the optimisation problem corresponding to different initial guesses. A common feature among the profiles of objective function is the abrupt variation in the values of the objective function and later, the trend is gradually smoothened and eventually nearly flattened during the allowed budget of iterations for solution estimation. This shows that step-size taken by the solver update the initial guess that improves the objective significantly and later, gradual improvement in the objective function is made corresponding to the update in the estimated solution. The objective function profiles against the iterations reveal that solver has explored the design space of the input variables for the solution estimation with sufficient iterations budget and is converged to local solution corresponding to different initial guesses. Moreover, feasible solutions are obtained corresponding to initial guesses, indicating the sufficient improvement in the objective function for the solution estimation of multi-objective function corresponding to three cases of power generation.

To investigate the impact of uncertain input conditions on the estimated solutions, 10000 uniformly distributed noise observations are generated for constructing Monte Carlo simulations and the variance in the estimated solutions is computed by Eqs. (14), (15). Fig. 7(b) confirms the bound applied on Power corresponding to three cases of power generation, i.e., [350, 360] for case 1, [490 MW, 500 MW] for case 2, and [650MW, 660 MW] for case 3, respectively. The estimated solution for the optimisation problem guarantees to have found feasible solution within the power generation interval of three cases. The variation in turbine heat rate and thermal efficiency is noted to be significant in the beginning during the estimation of optimal solution corresponding to the initial guesses and the improvement in the solution is smoothened gradually for the two performance parameters. This is attributed to the working mechanism of the solver to estimate the feasible solution on the face of the constraints incorporated in the optimisation problem. Monte Carlo simulation based variance computed for power is 0.001, 0.007 and 0.002 corresponding to estimated solution of 360 $\pm$ 11 MW, 496 $\pm$ 14 MW and 657 $\pm$ 15 MW, respectively. Referring to Fig. 7(c), the optimal value estimated for turbine heat rate is 7976 $\pm$ 122 kJ/kWh, 7487 $\pm$ 129 kJ/kWh and 7641 $\pm$ 118 kJ/kWh with the solver-driven variance of $\pm$ 8.74, $\pm$ 0.00031 and $\pm$ 0.0056, and Monte Carlo simulation based variance of 0.001, 0.003 and 0.002, respectively on the converged solutions corresponding to three cases, respectively. The optimal value of thermal efficiency is found to be 38.78 % with asymmetric confidence interval of [34.9, 39.3], 42.17 $\pm$ 1.06 % and 41.81 $\pm$ 0.88 % for solver-based variance of $\pm$ 0.063, $\pm$ 0.0000023 and $\pm$ 0.074 and Monte Carlo experiments driven variance of 0.021, 0.001 and 0.002, respectively on solution convergence for three modes of power generation from the power plant as depicted on Fig. 7(d).

The surface plot is constructed against three performance parameters, i.e., power, turbine heat rate and thermal efficiency, as presented on Fig. 7(e), which confirms the nonconvex function space for the multi-objective optimisation problem and the estimated feasible solutions corresponding to the three cases are mapped. We have demonstrated the effective integration of KAN model into the nonlinear programming framework for multi-objective optimisation of thermal power plant's operation under different power generation scenarios. It is anticipated

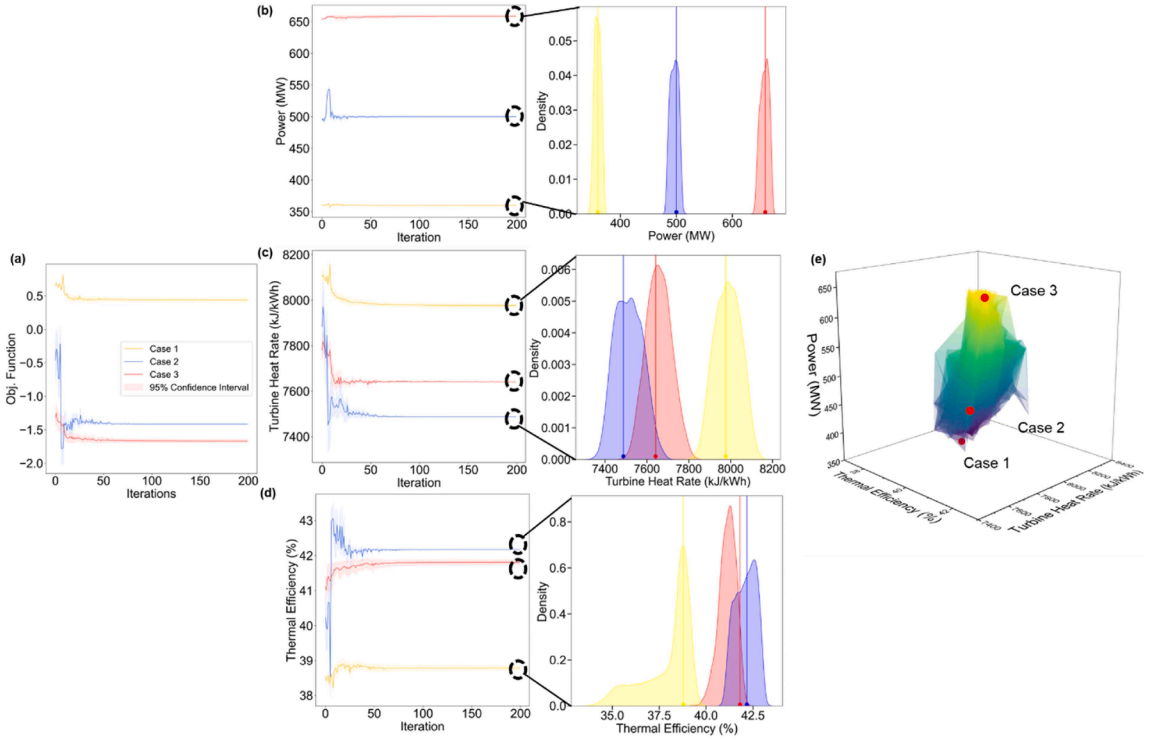



**Fig. 7.** KAN based multi-objective optimisation problem solved under three cases of power generation. (a) The progress of solver to estimate the solution in the iteration budget is depicted. The iterative solution convergence corresponding to initial guesses for (b) Power, (c) turbine heat rate, and (d) thermal efficiency is represented. (e) The optimal solutions are mapped on the surface plot corresponding to three cases of power generation.

that KAN based data analytics can increase the performance enhancement of energy systems including thermal power systems that increases the energy efficiency of power generation operation of thermal power plant. This research demonstrates the potential of KAN model for operation excellence of industrial-scale thermal power plant that enhances the operation quality of the power system which may significantly reduce the emissions discharge through optimal fuel consumption to support industrial decarbonisation of power sector.

## Conclusions

KAN algorithm offers improved interpretability performance than traditional MLP algorithm in low-dimensional and simulated data-driven studies. However, the systematic comparison of KAN with MLP for energy systems modelling and optimisation, particularly for industrial-scale power system may provide the insights about the suitability of the KAN and the proposed variants, as proposed in this paper, for the modelling tasks. In this paper, the algorithmic configuration of KAN and MLP are modified to embed with PCC correlation coefficient. The PCC based information quantification and embedding into the loss function tunes the parameters of KAN_PCC and MLP_PCC. The key conclusions of this research paper are as follows:

- KAN and MLP as well as KAN_PCC and MLP_PCC are applied on the two case studies of energy systems namely energy efficiency cooling (ENC) and energy efficiency heating (ENH) as well as performance parameters of 660 MW thermal power plants, i.e., power, turbine heat rate and thermal efficiency.
- KAN_PCC does not marginally improve the modelling performance on the test dataset for ENC, ENH, and power, turbine heat rate and thermal efficiency of thermal power plant. Although, KAN seems to provide nearly the same or slightly improved performance, in comparison with KAN_PCC, on the test datasets for the performance parameters of the two case studies. However, incorporating PCC in the architecture of MLP seems to have positive impact on the modelling performance of the algorithm for the considered case studies.
- The order of feature importance for the plant operation of thermal power plant integrating KAN or KAN_PCC with the SHAP based analysis framework is established. FWT, TAFR and MSP are found to be the most significant variables for KAN based predictions for power, turbine heat rate and thermal efficiency.
- KAN or KAN_PCC based models for power plant's performance parameters are incorporated in the optimisation framework of nonlinear programming technique for multi-objective optimisation of thermal power plant. The feasible optimal solutions are obtained, estimating maximum thermal efficiency of 42.17 $\pm$ 0.88 % and minimum turbine heat rate of 7487 $\pm$ 129 kJ/kWh corresponding to 500 $\pm$ 14 MW generation capacity of power plant.

This research papers highlights the impact of integrating the prior data information on the modelling performance of KAN and MLP for energy systems. The interpretable feature of KAN and its successful integration into the SHAP and nonlinear programming framework may accelerate the adoption of the KAN model for system analysis and optimisation of energy systems to support net-zero goal.

## Future work

The future work may systematically compare the compute efficiency and improvement in the modelling performance through KAN and MLP that may provide fundamental insights about the feasibility of the two algorithms for their potential applications for the analysis of energy networks and energy systems. Integrating the KAN architecture within the existing optimisation framework can be a technical challenge considering the nature of optimisation solvers and their compliance for the local or global optimisation of the system.

## CRediT authorship contribution statement

**Talha Ansar:** Writing – original draft, Software, Investigation, Formal analysis. **Waqar Muhammad Ashraf:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Data curation, Conceptualization.

## Declaration of competing interest

The authors declare no conflict of interest reported in this study.

## Acknowledgments

Waqar Muhammad Ashraf acknowledges that this work was supported by The Alan Turing Institute's Enrichment scheme.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.egyai.2025.100473.

## Data availability

The code written and data used to carry out this research is available at https://github.com/Waqar9871/KAN-vs-MLP/tree/main/Github.

## References

[1] Del Gaudio BL, et al. How do mobile, internet and ICT diffusion affect the banking industry? An empirical analysis. Eur Manag J 2021;39(3):327–32.
[2] Chatterjee A. Financial inclusion, information and communication technology diffusion, and economic growth: a panel data analysis. Inf Technol Dev 2020;26(3):607–35.
[3] Faheem M, et al. Smart grid communication and information technologies in the perspective of industry 4.0: opportunities and challenges. Comput Sci Rev 2018;30:1–30.
[4] Jin X, Yu W. Information and communication technology and carbon emissions in China: the rebound effect of energy intensive industry. Sustain Prod Consum 2022;32:731–42.
[5] Duan HK, et al. Enhancing the government accounting information systems using social media information: an application of text mining and machine learning. Int J Account Inf Syst 2023;48:100600.
[6] Chaudhuri R, et al. Adoption of robust business analytics for product innovation and organizational performance: the mediating role of organizational data-driven culture. Ann Oper Res 2024;339(3):1757–91.
[7] Sun Y, et al. CS2P: improving video bitrate selection and adaptation with data-driven throughput prediction. In: Proceedings of the ACM SIGCOMM conference; 2016.
[8] Mousavi S, et al. Data-driven prediction and optimization toward net-zero and positive-energy buildings: a systematic review. Build Environ 2023;242:110578.
[9] Ma S, et al. Data-driven sustainable intelligent manufacturing based on demand response for energy-intensive industries. J Clean Prod 2020;274:123155.
[10] Muhammad Ashraf W, Dua V. Driving towards net-zero from the energy sector: leveraging machine intelligence for robust optimization of coal and combined cycle gas power stations. Energy Convers Manag 2024;314:118645.
[11] Ashraf WM, Dua V. Artificial intelligence driven smart operation of large industrial complexes supporting the net-zero goal: coal power plants. Digit Chem Eng 2023;8:100119.
[12] Turja AI, et al. Waste heat recuperation in advanced supercritical CO2 power cycles with organic rankine cycle integration & optimization using machine learning methods. Int J Thermofluids 2024;22:100612.
[13] Khan Y, et al. Advanced cascaded recompression absorption system equipped with ejector and vapor-injection enhanced vapor compression refrigeration system: ANN based multi-objective optimization. Therm Sci Eng Prog 2024;49:102485.
[14] Hossain MM, et al. Numerical Investigation of a modified Kalina cycle system for high-temperature application and genetic algorithm based optimization of the multi-phase expander's inlet condition. Energy AI 2021;6:100117.
[15] Turja AI, et al. Multi-objective performance optimization & thermodynamic analysis of solar powered supercritical CO2 power cycles using machine learning methods & genetic algorithm. Energy AI 2024;15:100327.
[16] Baigh TA, et al. Enhancing thermodynamic performance with an advanced combined power and refrigeration cycle with dual LNG cold energy utilization. Heliyon 2024;10(15):e35748.

[17] Ashraf WM, Dua V. Data Information integrated Neural network (DINN) algorithm for modelling and interpretation performance analysis for energy systems. Energy AI 2024;16:100363.
[18] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw 1989;2(5):359–66.
[19] Leshno M, et al. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. Neural Netw 1993;6(6):861–7.
[20] Csáji BC. Approximation with artificial neural networks, 24. Hungary: Faculty of Sciences, Etvs Lornd University; 2001. p. 7.
[21] Hornik K. Approximation capabilities of multilayer feedforward networks. Neural Netw 1991;4(2):251–7.
[22] Ying X. An overview of overfitting and its solutions. In: Proceedings of the journal of physics: conference series. IOP Publishing; 2019.
[23] Wang Q, et al. Atmseer: increasing transparency and controllability in automated machine learning. In: Proceedings of the CHI conference on human factors in computing systems; 2019.
[24] Chattopadhyay A, et al. Neural network attributions: a causal perspective. In: Proceedings of the international conference on machine learning. PMLR; 2019.
[25] Zhang C, Zhang K, Li Y. A causal view on robustness of neural networks. Adv Neural Inf Process Syst 2020;33:289–301.
[26] Liu, Z., et al., KAN: Kolmogorov-Arnold networks. arXiv preprint arXiv: 2404.19756, 2024.
[27] Braun J, Griebel M. On a constructive proof of Kolmogorov's superposition theorem. Constr Approx 2009;30:653–75.
[28] Schmidt-Hieber J. The Kolmogorov–Arnold representation theorem revisited. Neural Netw 2021;137:119–26.
[29] Samadi, M.E., Y. Müller, and A. Schuppert, Smooth Kolmogorov Arnold networks enabling structural knowledge representation. arXiv preprint arXiv:2405.11318, 2024.
[30] Yu, R., W. Yu, and X. Wang, KAN or MLP: a fairer comparison. arXiv preprint arXiv: 2407.16674, 2024.
[31] Shukla K, et al. A comprehensive and FAIR comparison between MLP and KAN representations for differential equations and operator networks. Comput Methods Appl Mech Eng 2024;431:117290.
[32] Dong, C., L. Zheng, and W. Chen, Kolmogorov-Arnold Networks (KAN) for time series classification and robust analysis. arXiv preprint arXiv:2408.07314, 2024.
[33] Guo, C., et al., Physics-informed Kolmogorov-Arnold Network with Chebyshev polynomials for fluid mechanics. arXiv preprint arXiv:2411.04516, 2024.
[34] Mubarak AS, et al. Quasi-Newton optimised Kolmogorov-Arnold networks for wind farm power prediction. Heliyon 2024;10(23):e40799.
[35] Shen, H., et al., Reduced effectiveness of kolmogorov-arnold networks on functions with noise. arXiv preprint arXiv:2407.14882, 2024.
[36] Tsanas A, Xifara A. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy Build 2012; 49:560–7.
[37] Shapley, L.S., Notes on the n-person game—Ii: the value of an n-person game. 1951.
[38] Ashraf WM, et al. Optimization of a 660 MW e supercritical power plant performance—a case of industry 4.0 in the data-driven operational management part 1. Thermal efficiency. Energies 2020;13(21):5592.
[39] Ashraf WM, et al. Strategic-level performance enhancement of a 660 MWe supercritical power plant and emissions reduction by AI approach. Energy Convers Manag 2021;250:114913.
[40] Bergstra J, et al. Algorithms for hyper-parameter optimization. Adv Neural Inf Process Syst 2011;24.
[41] Bergstra J, Yamins D, Cox D. Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. In: Proceedings of the international conference on machine learning. PMLR; 2013.
[42] Riedmiller M, Lernen A. Multi layer perceptron. Machine learning lab special lecture, 24. University of Freiburg; 2014.
[43] Cuomo S, et al. Scientific machine learning through physics–informed neural networks: where we are and what's next. J Sci Comput 2022;92(3):88.
[44] Kingma, D.P., Adam: a method for stochastic optimization. arXiv preprint arXiv: 1412.6980, 2014.
[45] Dewolf N, Baets BD, Waegeman W. Valid prediction intervals for regression problems. Artif Intell Rev 2023;56(1):577–613.
[46] Vovk, V., A. Gammerman, and C. Saunders, Machine-learning applications of algorithmic randomness. 1999.
[47] Shapley LS. A value for n-person games. Contrib Theory Games 1953;2.
[48] Uddin GM, et al. Monte Carlo study of the molecular beam epitaxy process for manufacturing magnesium oxide nano-scale films. IIE Trans 2015;47(2):125–40.
[49] Uddin GM, et al. Artificial intelligence-based Monte-Carlo numerical simulation of aerodynamics of tire grooves using computational fluid dynamics. AI EDAM 2019; 33(3):302–16.
[50] Ashraf WM, Dua V. Partial derivative-based dynamic sensitivity analysis expression for non-linear auto regressive with exogenous (NARX) modelcase studies on distillation columns and model's interpretation investigation. Chem Eng J Adv 2024;18:100605.
[51] Grabisch M. K-order additive discrete fuzzy measures and their representation. Fuzzy Sets Syst 1997;92(2):167–89.
[52] Kumar R, Singh AK. Chemical hardness-driven interpretable machine learning approach for rapid search of photocatalysts. NPJ Comput Mater 2021;7(1):197.
[53] Ashraf WM, et al. Artificial intelligence modeling-based optimization of an industrial-scale steam turbine for moving toward net-zero in the energy sector. ACS Omega 2023;8(24):21709–25.
[54] Floudas CA. Nonlinear and mixed-integer optimization: fundamentals and applications. Oxford University Press; 1995.
[55] Powell MJ. A direct search optimization method that models the objective and constraint functions by linear interpolation. Springer; 1994.
[56] Shafer G, Vovk V. A tutorial on conformal prediction. J Mach Learn Res 2008;9(3).