Enabling industrial decarbonization through data-metrics guided interpretable machine learning

Waqar Muhammad Ashraf[1,2], Vivek Dua[1], and Ramit Debnath[2,3,4+]

[1]University College London, London, WC1E 7JE, UK
[2]University of Cambridge, Cambridge, CB2 1PZ, UK
[3]Caltech, Pasadena, 91125, US
[4]Indian Statistical Institute, Kolkata, 700108, India
[+]Corresponding author: rd545@cam.ac.uk

Abstract

Achieving industrial decarbonization requires accurate system-level modeling for the analysis of industrial systems. First-principle models, based on assumptions, are unable to forecast the actual condition of large-scale industrial systems, whereas pure data-centric machine learning (ML) algorithms do not possess an interpretable concept. This highlights the significance of addressing this issue to fully harness AI's potential for industrial decarbonization. The interpretable ML models can be trained by integrating the system's information, as quantified by data metrics, into the algorithmic design that offers improved interpretability performance for industrial system design, operation, and control. In this article, we offer a framework on how an interpretable ML-powered lever can significantly enhance energy efficiency, reduce resource consumption, and reduce emissions through smart operation management of industrial systems, including hard-to-decarbonize industrial sectors, thereby facilitating industrial climate action.

Introduction

With the advent of a new data-driven industrial age, the global digital transformation of our society, also known as digitalization, is rapidly shaping our existing anthropogenic systems and processes. However, we have yet to fully understand the impact of this digitalization, raising open-ended questions about its potential to cause more harm to the environment, economy, and society than its perceived benefits [1, 2]. Big data, machine learning (ML), and artificial intelligence (AI) algorithms drive the digital transformation at a systems level, attempting to meaningfully interpret the vast amount of data generated from our economic activities. This is often referred to as the building blocks of 4th Industrial Revolution or Industry 4.0 [3, 4].

There is a direct link between economic activities and greenhouse gases (GHGs) emissions, driven by the production of goods and services in industries for society's consumption. Rapidly rising GHGs emissions are one of the major causes of climate change, and it is urgent that we reduce these emissions through various climate actions like mitigation, adaptation, and carbon dioxide removal. Therefore, industries are crucial leverage points for an emission reduction system change, and AI can help making these decisions [3, 5]. Research shows that industrial decarbonization using AI at large scale depends on

https://doi.org/10.33774/coe-2024-3r5v0 ORCID: https://orcid.org/0000-0003-0727-5683

the development of cyber-physical systems (CPS) [4]. CPS is an integration and intelligent orchestration of computation and physical processes where they operate on different spatial and temporal scales, exhibit multiple and distinct behavioral modalities, and interact with each other in ways that change with context [6]. CPS are technologically challenging and demand innovation in AI algorithm design that is interpretable and explainable. This is especially true when creating CPS for industrial processes, where data distribution becomes highly nonlinear and complex, and empirical models fail to explain this high-dimensionality system's operation [7]. At the same time, we argue that these systems must also be accessible and understandable to policymakers who will ultimately design appropriate decarbonization policies at local, regional and international level [8].

AI serves as a lever to drive the higher performance of the industrial systems through the optimal consumption of the resources to provide same energy services, reduction in GHGs intensity, and managing the demand and supply of the power grids through deep predictive capabilities. A recent research study finds that AI can curb down 2.4 Gt GHGs while adding $ 5.2 Trillion to the global economy relative to the baseline projections until 2030 [9]. Considering the enormous potential of AI towards the environmental sustainability, it becomes imperitive to understand the interpretable performance of the AI based algorithms. Theoretically, we rely on the following definitions of AI interpretability, proposed by Doshi-Velez and Kim *"the ability to explain or to present in understandable terms to a human"* [10] and Miller *"the degree to which a human can understand the cause of a decision"* [11].

There are three major frameworks for evaluating AI interpretability: application-grounded, human-grounded, and functionally-grounded evaluations [10, 12]. Application-grounded evaluation checks how the results of the interpretation process affect the human, domain expert, or end-user in terms of a specific and well-defined task or application. For example, it results in better identification of errors or less discrimination. In human-grounded evaluations, any human end-user can serve as the tester instead of a domain expert. The objective is not to assess an interpretation's suitability for a particular application, but to assess its quality in a broader context and gauge the accuracy of capturing general concepts. Functionally grounded evaluation does not require any experiments that involve humans but instead uses formal, well-defined mathematical definitions of interpretability to evaluate the quality of an interpretability method. For example, once a class of models passes some interpretability criteria through human- or application-grounded experiments, we can use mathematical definitions to further rank the interpretability models' quality [10].

In this perspective, we expand upon the theoretical framework of functionally-grounded AI interpretability, leveraging interpretable ML to train the models that are interpretable to facilitate the decision making that enhances the industrial performance and supports the industrial decarbonization. For example, we argue that while first-principle models attempt to interpret the physical phenomenon in nature, their accuracy and adaptability may be limited to real-life applications, particularly in complex and large-scale aging industrial systems where lack of system's knowledge and information are difficult to model and then incorporate in the physical models; thereby leading to the significant deviation in the first principle model-based simulated responses and actual behavior of the system. Since, data capture the system related information that is well-established in the data silos, ML models can effectively recognize the data patterns and thus, are customized to predict the performance indicator(s) of the industrial systems. Moreover, updating the ML model requires to carryout adaptive training of the model on the recent data of the application that is flexible and computationally inexpensive in comparison to parameters estimation for the update of the first-principle models.

We structure the paper as follows: in the next two sections we expand on the evolutionary paradigm of industrial data and how it relates to industrial decarbonization's objectives. Next, we present the existing challenges and opportunities of data-centric industrial decarbonization. After scoping the challenges and opportunities, we provide insights into fostering innovation in AI algorithm design through embedding the statistics metrics and aligning it with the objectives of industrial decarbonization. This includes investigating the structural comparison of cutting-edge neural network models such as physics-informed neural networks (PINNs) with the recently introduced data information integrated neural network (DINN). Finally, we present a way forward for supporting climate action via interpretable ML based design, operation and control of the industrial systems that offers techno-economic and environmental benefits.

Historical shift in human-learning paradigm to data-centricism

Human capacity for understanding the physical world has evolved through empirical, theoretical, and computational science approaches, leading us now into the era of data-driven science. Initially, empirical sciences attempted to explain an event without theoretical insights. The empirical sciences eventually evolved into the realms of theoretical sciences that lasted until around 1950, and the discoveries of universal laws enabled us to understand and explain the physical phenomena in nature. The theoretical sciences era gave way to the computational sciences approach, which combined both empirical and theoretical methods and persisted until roughly 2010. Since then, better computers and technology have created a new way of learning called "data-driven science". This way of learning focuses on getting knowledge from the data we collect at the points where different fields of knowledge meet [13]. We argue that this brings paradigm-shifting potential to decarbonize industrial systems.

Data driving the evolution of industrial systems decarbonization

Traditionally, experimentally validated mathematical models are created to explain the characteristics of engineering systems [14]. However, their accuracy and adaptability may be limited, particularly in complex and large-scale aging industrial systems that operate under synchronized and/or sequential processes. This is further highlighted conceptually in Figure 1(a) that as we observe increasingly larger data observations, the mathematical models become more and more complex to interpret the data occurrence. As the data-occurrence and distribution become highly nonlinear and complex, the mathematical models fail to explain the data generation process of the high-dimensional data. Applying this concept to the scale of industrial operations where hyperdimensional control space and data volumes are quite large, creating accurate mathematical models for the dynamic operation of large industrial systems is quite challenging [7]. Moreover, the simulation and optimisation analyses based on the large-sized mathematical models become computationally prohibitive that further limits the applicability of the mathematical models for the large scale industrial systems.

To address these challenges, we turn to the data-driven learning paradigm, which seeks to exploit the data to construct customized models of the system under consideration. The data-driven models excel at capturing nonlinear dynamics for predictive analytics, uncovering the underlying physics hidden under the hyper-dimensional variable space [15–19]. In recent years, ML has evolved from "shallow" application domains like recognizing cats on the internet to more complex constructs such as graphs and symbolic representations, invariances, and positional embeddings, which represent deeper levels of scientific abstraction. Currently, classical ML methods are making significant impacts in biological and medical fields [20–22], despite these fields often have poorly defined entities and governing laws to interpret system behavior.

In contrast, the physical sciences have rigorously defined entities and relationships based on conservation laws, invariances, and causal relationships [23]. Applying ML models without taking these constraints into account can result in inaccurate results if the model training violates relationships. Therefore, pure data-centric approaches face criticism for their lack of transparency and explainability, limiting their ability to capture real-world complexity. One powerful ML approach that recently emerged is symbolic regression. Methods like SISSO [24], SinDy [25], and PySR [26] derive symbolic expressions from observational data for applications such as predicting chemical reactivity, relating crystallographic structures to atomic radii, and defining electron transfer directions between elements. However, the interpretability of these expressions remains challenging for medium to large-scale problems, often missing the opportunity to address the model's interpretability problem.

Challenges and opportunities for data-driven models to support industrial decarbonization

While the potential of data-driven models for real-life problems is promising, we must address significant challenges to train a generalized and deployable ML model to support industrial decarbonization. Importantly, it is not just about collecting data and feeding it into an ML algorithm, as the adage "garbage in, garbage out" reminds us for data-driven modelling [27]. The data obtained from the experiments often contains observational biases and confounders, limiting its interpretability. Similarly, sensor-based measurements can have intermittent, freezing, and drift errors, potentially introducing noise into the recorded observations. Another challenge is defining the system boundaries and scope of the problem for the ML-based studies in terms of the selection of the variables. During training, a deep learning model minimizes its loss by absorbing spurious correlations, including confounding factors and selection biases that adversely affects the generalization capability of the model. These connections, which have nothing to do with real causes, can change between training and application domains, which makes it hard to do well with out-of-distribution data [28].

Some suggestions exist in the literature for the ML technique selection considering the nature of data-distribution, number of variables and suitability of the algorithm for the quantitative and qualitative nature of the problem [29–31], the selection of the ML technique for the data-driven modelling remains an open question. Another important challenge is to assess the uncertainty bounded with the point-predictions made by the ML model. The point-prediction enveloped uncertainty sources back to the modelling inaccuracy and the hyperdimensional parameters space associated with the model-structure [32]. The Bayesian and ensembles methods lack the validity issues of the constructed prediction intervals [33] while, transductive conformal prediction approach is computationally expensive to draw the prediction intervals and the model needs to be retrained for estimating the prediction intervals for the new input conditions [34].

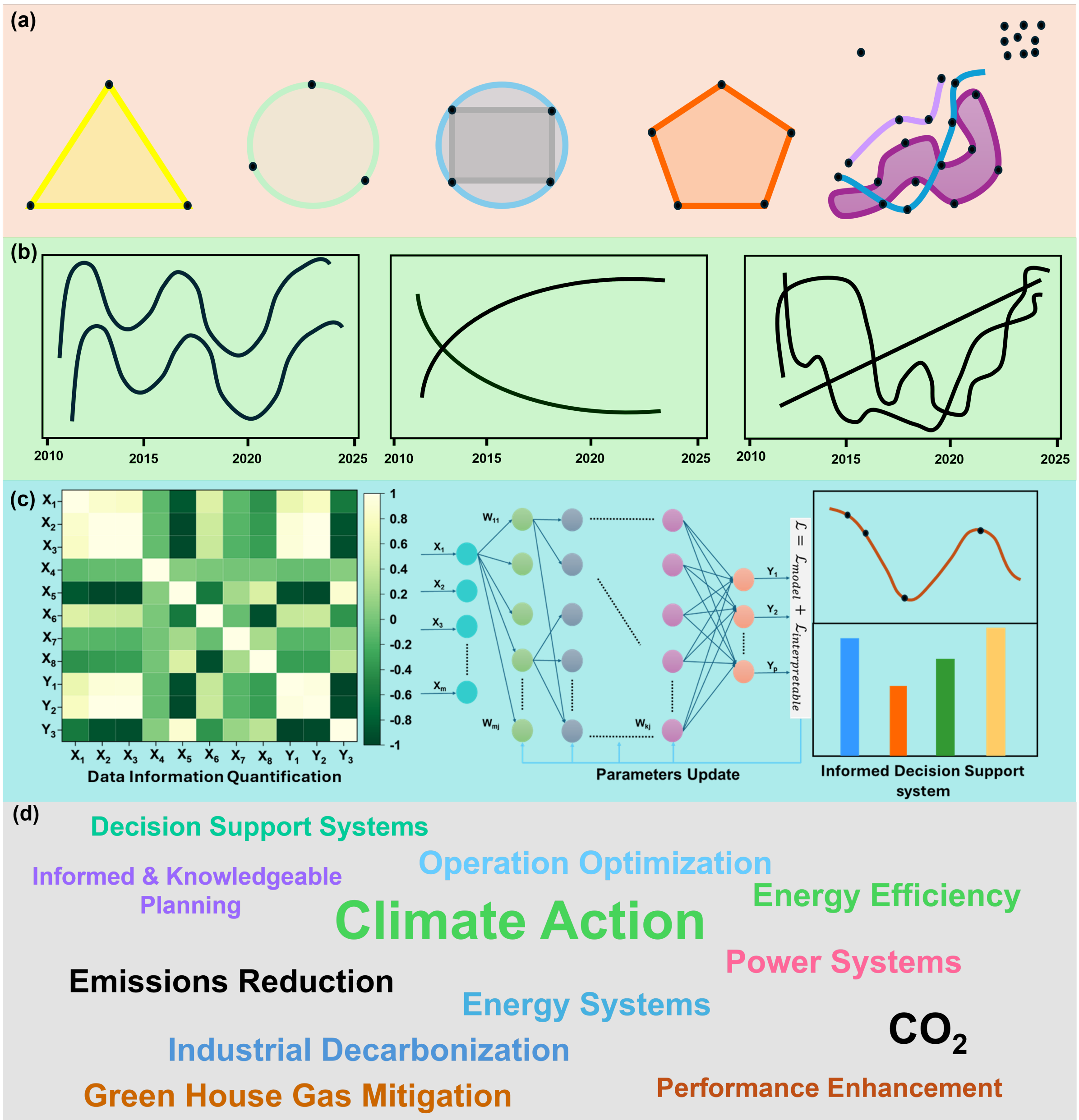

Figure 1: How data direct to update the model for explaining an event. (a) Observing small number of data observations results in fitting different low-dimensional mathematical models. As the size of randomly observed data increases, traditional mathematical models fail to explain the data-generation process. (b) Observing the long-term variables interactions for the time-delayed, monotonic and highly nonlinear relationships typically characterized for the industrial complexes. (c) The data-based information is extracted by the relevant statistical metrics to get an insight of the behaviour of the industrial systems. The loss function of the neural network comprises of two loss terms, (i) typical modelling-based loss, and (ii) loss occurred due to interpretation metric embedded in the loss function. (d) The data-metric(s) guided trained neural network can support climate action through optimal performance of industrial systems by mitigating the GHGs and enhancing the energy efficiency through model-based decision support system, and informed planning.

Explaining the model-based predictions, formally called explainable AI (XAI), is a hot research issue among the research communities. The XAI approach seeks to investigate the transparency, justification, informativeness and reliability for the model-based predictions [35]. The XAI techniques are categorized into feature attribution, instance based, graph convolution based, self-explaining and uncertainty estimation [36]. Counterfactual, attention mechanisms, human-readable rule extraction, adversarial explanation, transfer learning and integrating human feedback etc., are some of the XAI categories used for explaining the model-based predictions. These techniques, either ante-hoc or post-hoc, suffer owing to their working principles for explaining the model-based predictions for hyper dimensional parameter space embedded in the ML models. Researchers are also exploring active learning based on Gaussian processes to enhance the interpretability of the model. However, these methods face significant challenges when assuming a Gaussian error distribution, which limits their applicability to complex real-world datasets. However, the data-centric, interpretable modeling approach offers great potential to efficiently utilize system-specific information for the development of the ML model.

Despite the aforementioned challenges, as summarized in Table 1, that still exist towards the artificial general intelligence, the ML models, trained under extensive preprocessing, feature engineering, and rigorous hyperparameters tuning, have demonstrated practical utility for applications such as power systems [37–42] and industrial cooling and heating systems [43, 44]. The predictive accuracy of the ML models is improved under adaptive training as the data generation pattern and process change from the practical application system. A well-predictive ML model enables to develop informed operating strategies and decision support systems, thereby enhances the energy efficiency of the industrial systems. As the industrial systems become more and more digital, data-driven sciences are used to improve the design, operation, and control of real-life industrial systems. The data-driven sciences can estimate the robust and energy-efficient operating conditions for the existing in-operation industrial systems to reduce the emissions discharge without the significant retrofitting solutions [37, 45] which helps the industry act on climate change on the face of economic and societal challenges.

Table 1: Challenges that exist for the data-driven analytics integrating ML models and optimisation techniques for the performance enhancement of industrial systems.

Scope	Challenges	Mitigation
Data Curation & Access	• Manual data collection from the on-site systems • Temporal inconsistencies of data-recording for multi-level operation • Inadequate data recording and storage facilities • Human error and bias in the measurement • Data loss • No access to data storage or partial data retrieval at improper data sampling rate	• Digitalization of the data recording • Developing strategies to reduce the data sampling frequencies through advanced metering and measurement systems • Implementing secure data storage systems and maintaining the data-logs of the industrial systems • Integrating advanced supervisory information system for data extraction at the desirable sampling rate
Data processing	• Data distribution is skewed and imbalanced • Sensor-measurements induced errors • Feature selection for the data-driven modeling • Data scaling for the modeling tasks	• Deploying techniques to sample the data from big-clusters for balancing the data-distribution across the operating ranges • Applying outliers' removal techniques and visualize the data to identify the potential outliers [40] • Consulting with domain-experts and applying statistical techniques for feature selection and dimensionality reduction

Continued on next page

Table 1 – continued from previous page

Scope	Challenges	Mitigation
ML model development	• Selection of the ML model corresponding to the operating level (component, system, strategic) of the industrial systems • Rigorous model evaluation through robust metrics on the unseen data (data sample from the industrial operation) • Estimating the uncertainty bounded with the model prediction	• Neural network-based algorithms for the quantitative nature of the data and classification-based algorithms for the qualitative nature of the dataset [46] • Apply prediction interval estimation techniques (conformal prediction, direct interval estimation) for the uncertainty quantification [32, 47]
Model based optimisation under uncertainty	• The non-convex optimisation problem integrating the ML models poses computational issues to estimate a solution • The solver(s) may converge to a feasible solution that may not align well with the domain-knowledge • Introducing uncertainty and its evaluation for model-based optimisation is challenging and computationally expensive	• Introducing data information-based constraints in the optimisation problem that guide the solver(s) to estimate an efficient solution • Development of scalable and computationally effective tools for ML-model based optimisation under uncertainty for large-scale industrial data problems

Continued on next page

Table 1 – continued from previous page

Scope	Challenges	Mitigation
Model calibration and security	• Consistent model-calibration requires experts for in-house analysis • Cyber-attacks and model security • Self model adaptation and tuning	• Robust algorithm design for the self-model calibration with human-in-the-loop to monitor the model deterioration and accuracy • Secure model deployment environment against cyber attacks

Algorithm design innovation aligned with industrial decarbonization

The need for innovations in algorithm design to train reliable models for industrial systems is clear. Information and communication technologies record vast amounts of real-time data from industrial operations. These datasets possess information about system dynamics accumulated over years of industrial system's operation as shown graphically on Figure1(b). Notably, the datasets capture the interactions and nonlinear dynamics at the intersection of interdisciplinary knowledge domains like chemical and/or bio-chemical processes, material science, and electro-mechanical characterization that essentially represent the "Physics" of the "specific" system [14]. This provides an opportunity to incorporate "system physics" into the algorithmic design, similar to how Physics-Informed Neural Networks (PINNs) guide model parameters' updates while respecting system physics to train the model. Thus, there is a need to utilize and/or innovate statistical terms that accurately quantify the strength of the features-relationships as established in the big data. The statistical metrics commonly used to quantify the association between the two variables are mentioned in Table 2. Considering the nature of variable (continuous, categorial, mixed), the statistical metric(s) relevant to available information and type of the problem is selected and can be deployed to develop interpretable ML model [15].

Methodologically, there is an opportunity to leverage the 'physics' that exists in the features-relationships in the form of the statistical data-metric quantifying the association between the variables. Similar to PINNs, ML models can integrate these statistical terms into the loss function (L) that contributes to parameter update during the model training. The customized L integrating the Pearson correlation coefficient, and the parameters update by gradient descent with momentum algorithm is given as [15]:

$$L = a_1 \cdot \frac{\sum(D - Z)^2}{N} + a_2 \cdot \frac{\sum_{i=1}^{m}(r_{X_i|D} - r_{X_i|Z})^2}{N} \tag{1}$$

$$W_1^{\text{new}} = W_1 - \eta V_{W_1} \tag{2}$$

$$V_{W_1} = \beta V_{W_1} + (1 - \beta)\frac{\partial L}{\partial W_1} \tag{3}$$

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial Z}\frac{\partial Z}{\partial p_2}\frac{\partial p_2}{\partial y_1}\frac{\partial y_1}{\partial p_1}\frac{\partial p_1}{\partial W_1} \tag{4}$$

here, D and Z are the true and model-simulated response, N is the total data observations, m is the total number of input variables, $r_{X_i|D}$ and $r_{X_i|Z}$ are the Pearson correlation coefficient computed for X_i with respect to D and Z respectively. a_1 and a_2 are the weightage factors; W_1 is matrix containing the weight connections from input to hidden layer of data information integrated neural network (DINN) model, V_{W_1} is the velocity matrix defined on W_1; η and β are the learning rate and momentum parameters respectively; and $\frac{\partial L}{\partial W_1}$ is the partial derivative of L with respect to W_1 that is calculated by eq. (4) and contributes to weight update (W_1^{new}). The rest of model parameters (W_2, b_1, b_2) are also updated on the customized L during the iterative model training. The graphical representation of the information flow and processing in DINN is presented on Figure 2.

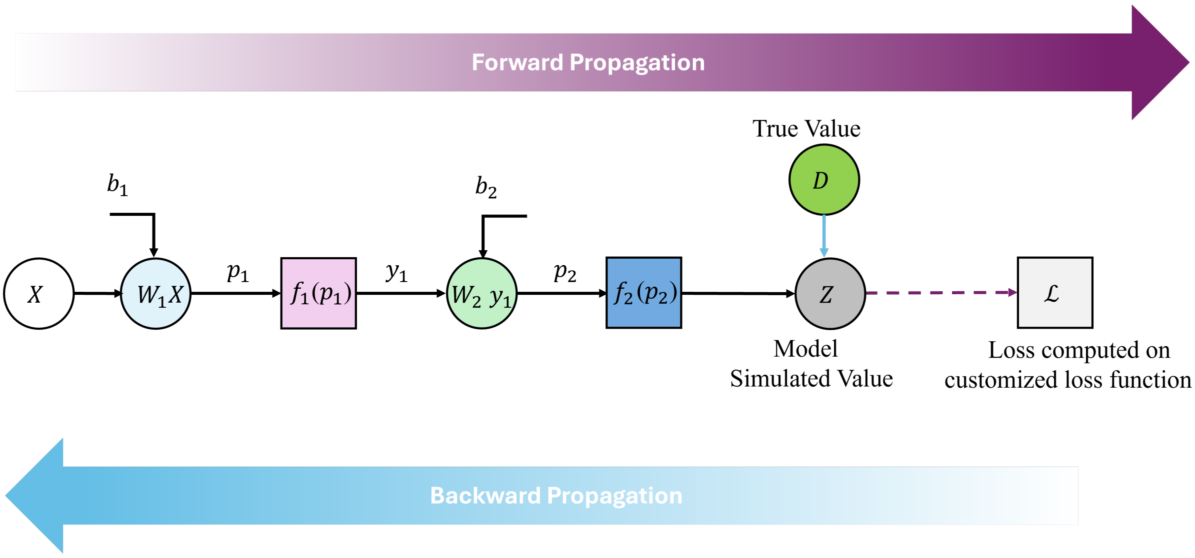

Figure 2: The information flow in the architecture of data information integrated neural network (DINN) algorithm. The parameters are updated iteratively on the customized loss function

Table 2: Leveraging the association of the variables (continuous, categorical, mixed) for the development of interpretable ML model.

Nature of Variables	**Metrics**	**Function**
Continuous	Pearson Correlation Coefficient (r): $r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}$ x and y are two variables with the mean of $\bar{x}$ and $\bar{y}$ respectively	• Measures the linear relationship between the two variables • Ranges from -1 to +1

Continued on next page

Table 2 – continued from previous page

Nature of Variables	Metrics	Function
	Spearman's Rank Correlation Coefficient (ρ): $$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$ d is the deviation between the ranks of two variables for each data pair; n is the total number of data points	• Measures the strength and direction of the monotonic relationship between two ranked (ordinal) variables • Ranges from -1 to 1
	Coefficient of determination (R^2): $$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$	• Represents the proportion of the variance in the dependent variable that is predictable from the independent variable • Ranges from 0 to 1
	Kendall's Tau (τ): $$\tau = \frac{C - D}{C + D}$$ C and D are number of concordant and discordant pairs	• Measures the strength and direction of the association between two ranked (ordinal) variables • Ranges from -1 to 1
Categorical	Chi-Squared (χ^2): $$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$ O and E represent the observed and expected frequencies to observe the values of the categorical variable	• Provides a p-value indicating the significance of the association

Continued on next page

https://doi.org/10.33774/coe-2024-3r5v0 **ORCID:** https://orcid.org/0000-0003-0727-5683

Table 2 – continued from previous page

Nature of Variables	Metrics	Function
	Cramér's V: $$\sqrt{\frac{\chi^2}{n(k-1)}}$$ n is the number of observations and k is the smaller of the number of levels of the two variables	• Measures the strength of association between two categorical variables • Ranges from 0 to 1
	Goodman and Kruskal's lambda (λ): $$\lambda = \frac{E_1 - E_2}{E_1}$$ • E_1 is the overall non-modal frequency • E_2 is the sum of the non-modal frequencies for each value of the independent variable	• Measures the strength of association between two categorical variables, indicating the proportion of error reduced in predicting one variable using the other • Ranges from 0 to 1
	Contingency Coefficient (C): $$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$ n is the number of points in the dataset	• Measures the degree of association between two categorical variables • Ranges from 0 to a value less than 1

Continued on next page

Table 2 – continued from previous page

Nature of Variables	Metrics	Function
Mixed	Point-biserial Correlation Coefficient (r_{pb}): $$r_{pb} = \frac{\bar{X}_1 - \bar{X}_0}{s_X}\sqrt{\frac{n_1 n_0}{n(n-1)}}$$ where, • $\bar{X}_1$ is the mean of the continuous variable for the group where the binary variable is 1 • $\bar{X}_0$ is the mean of the continuous variable for the group where the binary variable is 0 • s_X is the standard deviation of the continuous variable • n_1 is the number of observations where the binary variable is 1 • n_0 is the number of observations where the binary variable is 0 • n is the total number of observations	• Measures the strength and direction of association between a continuous variable and a binary variable • Ranges from -1 to 1
	Eta-squared (η^2): $$\eta^2 = \frac{SS_{effect}}{SS_{total}}$$ where, • SS_{effect} is the sum of squares for the effect • SS_{total} is the total sum of squares for all effects, errors and interactions	• Measures the proportion of variance in a continuous dependent variable that is associated with one or more categorical independent variables • Ranges from 0 to 1

Way forward: supporting climate action via interpretable ML

In the previous sections, we have extensively explained the challenges in training a well-predictive and robust ML model for the modelling and optimisation of the industrial systems. We have also included the key data metrics that quantify the features association out of the data sourced from the considered application. Researchers can investigate the integration of relevant and intuitive (compliant with domain expertise) statistical terms into L, system-specific modeling constraints, and customized stopping criteria to train robust and generalized ML model. Recently, a research study reports the integration of the Pearson correlation coefficient into the L that provides the improved interpretable performance of the DINN model compared with those of a standard artificial neural network [15]. The interpretable ML models have the potential to restore the trust of the industrial community into ML based analytics, and contributing to develop robust decision support systems to support the decision making and operation excellence of industrial sectors that are even hard to decarbonize like steel, chemical, cement, shipping and aviation.

AI serves as a lever to support the climate action through energy efficiency improvement, emissions reduction, and performance enhancement of the industrial as highlighted on Figure 1(c,d). To unlock the full potential of AI towards the environmental sustainability, five principles are the 'enablers' for this data-sciences paradigms that include (i) raising awareness, value alignment and fostering interdisciplinary collaboration, (ii) performing robust AI-powered data-analytics that cater economic, societal and environmental concerns, (iii) building digital infrastructure and data-access, (iv) upskilling and reskilling for the adapting to sectoral adaptations, and (v) encouraging R&D from research to commercial deployment. The value of AI sector is well received by the governments and industrial stakeholders, it is of high need to innovate reliable, interpretable and efficient ML based algorithms that facilitate the decisions at the system and plenary boundaries to pave the way for deep industrial decarbonization.

ML based design, operation and control of industrial systems support the climate action

The industrial systems are essentially synchronized operation of the industrial processes that are maintained by electro-mechanical devices including pumps, compressors, conveyor belts etc. The design engineer can deploy the design data and constraints for the training of ML model and/or hybrid modelling approach to expedite the component and process design [48, 49] considering the nonlinear interactions that are well-captured by the ML model. The optimal component and system design offers the higher energy efficiency and efficient resource utilization to reduce the material consumption and carbon footprint to support the climate action.

Recently, ML is being used for the system-level modelling of the industrial systems including process industries [50, 51], power [52, 53] and transportation systems [54]. While the ML models can effectively predict the state of the operation based on the operating conditions and fault detection, the models can estimate the optimized operating conditions for the operation of the industrial systems [45]. The operation of the industrial systems can be analysed with the view of minimization of resource consumption, multi-objective optimisation of the techno-economic and environmental concerns, and the availability of the system. Moreover, the decision support system and operation management strategies guided by ML models can support the optimal performance of the industrial systems on the face of emissions constraints and product demand to facilitate the industrial decarbonization.

The data-driven control [55] and learning from the data [56] are the promising domains in the control systems that leverage the data-driven models for the control of the systems. While, the problem-specific data-metrics integrated into the algorithm design has the potential to train an interpretable ML model that respects the system's constraints, the trained ML model can be embedded into the control systems for the control applications [57]. The promising utility of ML based control is to forecast the optimal control trajectories for the dynamic operation of the industrial systems that the control system should follow under the human supervision to realize the optimal control of the industrial systems. The ML model-based control systems can estimate the energy-efficient control actions to limit the emissions discharge in the allowable limits during the operation of the industrial systems thereby curbing down the emission volumes to support the climate action.

While AI based algorithms can mine the complex patterns hidden in the high dimensional and hyper volume datasets, engineers and industrial managers can leverage the AI potential for energy-efficient and cost-effective system design, operation and control of the industrial systems. The effective system modelling and representation through interpretable ML algorithm lay the foundation for the construction of digital twin that should be both robust and resilient, as well as transform and renew timely in response to adverse events to facilitate the decision making for the technical, economical and environmental management towards the climate action.

Concluding Remarks

As we are living into fourth learning paradigm, i.e., data-driven sciences, it becomes crucial to extend the reliable application of machine intelligence for the modelling and optimisation of industrial systems to support the industrial decarbonization. The ML based models offer flexible approach to model the large-scale operation of industrial systems, as opposed to first-principle modelling, and the models can be effectively adapted to predict the state of the industrial systems.

The industrial systems store the heap of data on the years of operation and the representative association between the variables is almost established. Thus, the features association as computed on the data-metrics represent the system-specific physics that can be introduced in the algorithm like PINN works. The integration of the data metrics into the loss function of the ML algorithms can well update the model parameters and the trained model offers improved interpretability performance. As a result, the interpretable ML model can be deployed for the optimal design, operation and control of the industrial systems to enhance the system performance from the dimension of energy efficiency improvement, emissions reduction, optimal resource utilization to developing robust decision support system and effective operation management for the industrial systems.

At one side, AI drives the higher performance of the industrial systems, it also contributes to the GHGs reduction (2.4 Gt) and value addition in the economy ($ 5.2 Trillion by 2030) [9]. Therefore, it is anticipated that the ML based industrial analytics backed by domain expertise and exploiting the rich industrial data can accelerate for the robust and smart operation of the industrial systems that enables the sustainable future against the climate crisis.

Acknowledgement

WMA acknowledges to have received the funding (CMMS-PHD-2021-006) from The Punjab Education Endowment Fund (PEEF) to pursue PhD at University College London. RD acknowledges the support from the Bill and Melinda Gates Foundation [OPP1144], Cambridge Humanities Research Grants (2023) and UKRI Responsible AI IO0008 Grant Ref: EP/Y009800/1.

Author contributions

W.M.A. initiated the project and convened the team. W.M.A., V.D and R.D. conceptualized the study. W.M.A wrote the first draft of the manuscript with inputs from R.D. and V.D. All authors read, revised and approved the manuscript.

Competing Interests

The authors declare no competing interests to disclose in carrying out this study.

References

1. Creutzig, F. *et al.* Digitalization and the Anthropocene. *Annual review of environment and resources* **47,** 479–509 (2022).
2. Debnath, R. *et al.* Harnessing human and machine intelligence for planetary-level climate action. *npj Climate Action* **2,** 20 (2023).
3. Coyle, D. & Weller, A. "Explaining" machine learning reveals policy challenges. *Science* **368,** 1433–1434 (2020).
4. Inderwildi, O. & Kraft, M. *Intelligent Decarbonisation: Can Artificial Intelligence and Cyber-Physical Systems Help Achieve Climate Mitigation Targets?* (Springer Nature, 2022).
5. Rolnick, D. *et al.* Tackling climate change with machine learning. *ACM Computing Surveys (CSUR)* **55,** 1–96 (2022).
6. Lee, E. A. *Cyber physical systems: Design challenges* in *2008 11th IEEE international symposium on object and component-oriented real-time distributed computing (ISORC)* (2008), 363–369.
7. Bhutani, N., Rangaiah, G. & Ray, A. First-principles, data-based, and hybrid modeling and optimization of an industrial hydrocracking unit. *Industrial & engineering chemistry research* **45,** 7807–7816 (2006).
8. Inderwildi, O., Zhang, C. & Kraft, M. in *Intelligent Decarbonisation: Can Artificial Intelligence and Cyber-Physical Systems Help Achieve Climate Mitigation Targets?* 17–28 (Springer, 2022).
9. Joppa, L. & Herweijer, C. *How AI can enable a Sustainable Future* Technical Report (Microsoft, PwC UK, 2019). `https://news.microsoft.com/wp-content/uploads/prod/sites/53/2019/04/PwC-Executive-Summary.pdf`.

10. Doshi-Velez, F. & Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608* (2017).

11. Miller, T. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* **267,** 1–38 (2019).

12. Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. Explainable ai: A review of machine learning interpretability methods. *Entropy* **23,** 18 (2020).

13. Karagiorgi, G. *et al.* Machine learning in the search for new fundamental physics. *Nature Reviews Physics* **4,** 399–412 (2022).

14. Pantelides, C. C. & Renfro, J. G. The online use of first-principles models in process operations: Review, current status and future needs. *Computers & Chemical Engineering* **51,** 136–148 (2013).

15. Ashraf, W. M. & Dua, V. Data Information integrated Neural Network (DINN) algorithm for modelling and interpretation performance analysis for energy systems. *Energy and AI* **16,** 100363 (2024).

16. Uddin, G. M. *et al.* Artificial intelligence-based Monte-Carlo numerical simulation of aerodynamics of tire grooves using computational fluid dynamics. *Ai Edam* **33,** 302–316 (2019).

17. Tariq, R., Abatal, M. & Bassam, A. Computational intelligence for empirical modeling and optimization of methylene blue adsorption phenomena using available local zeolites and clay of Morocco. *Journal of Cleaner Production* **370,** 133517 (2022).

18. Tariq, R. *et al.* Deep learning artificial intelligence framework for sustainable desiccant air conditioning system: Optimization towards reduction in water footprints. *International Communications in Heat and Mass Transfer* **140,** 106538 (2023).

19. Ashraf, W. M. *et al.* Artificial intelligence based operational strategy development and implementation for vibration reduction of a supercritical steam turbine shaft bearing. *Alexandria Engineering Journal* **61,** 1864–1880 (2022).

20. Libbrecht, M. W. & Noble, W. S. Machine learning applications in genetics and genomics. *Nature Reviews Genetics* **16,** 321–332 (2015).

21. Angermueller, C. *et al.* Deep learning for computational biology. *Molecular systems biology* **12,** 878 (2016).

22. Shehab, M. *et al.* Machine learning in medical applications: A review of state-of-the-art methods. *Computers in Biology and Medicine* **145,** 105458 (2022).

23. Kalinin, S. V. *et al.* Physics is the new data. *arXiv preprint arXiv:2204.05095* (2022).

24. Ouyang, R. *et al.* SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Physical Review Materials* **2,** 083802 (2018).

25. Kaptanoglu, A. A. *et al.* Pysindy: A comprehensive python package for robust sparse system identification. *arXiv preprint arXiv:2111.08481* (2021).

26. Cranmer, M. *et al.* Discovering symbolic models from deep learning with inductive biases. *Advances in neural information processing systems* **33,** 17429–17442 (2020).

27. Geiger, R. S. *et al. Garbage in, garbage out? Do machine learning application papers in social computing report where human-labeled training data comes from?* in *Proceedings of the 2020 conference on fairness, accountability, and transparency* (2020), 325–336.

28. Nagarajan, V., Andreassen, A. & Neyshabur, B. Understanding the failure modes of out-of-distribution generalization. *arXiv preprint arXiv:2010.15775* (2020).

29. Raschka, S. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808* (2018).

30. Doan, T. & Kalita, J. *Selecting machine learning algorithms using regression models* in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)* (2015), 1498–1505.

31. Mahesh, B. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]* **9,** 381–386 (2020).

32. Shafer, G. & Vovk, V. A tutorial on conformal prediction. *Journal of Machine Learning Research* **9** (2008).

33. Dewolf, N., Baets, B. D. & Waegeman, W. Valid prediction intervals for regression problems. *Artificial Intelligence Review* **56,** 577–613 (2023).

34. Saunders, C., Gammerman, A. & Vovk, V. Transduction with confidence and credibility (1999).

35. Haque, A. B., Islam, A. N. & Mikalef, P. Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. *Technological Forecasting and Social Change* **186,** 122120 (2023).

36. Islam, M. U. *et al.* The past, present, and prospective future of xai: A comprehensive review. *Explainable Artificial Intelligence for Cyber Security: Next Generation Artificial Intelligence,* 1–29 (2022).

37. Ashraf, W. M. & Dua, V. Artificial intelligence driven smart operation of large industrial complexes supporting the net-zero goal: Coal power plants. *Digital Chemical Engineering* **8,** 100119 (2023).

38. Ashraf, W. M. *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* **13,** 5592 (2020).

39. Ashraf, W. M. *et al.* Strategic-level performance enhancement of a 660 MWe supercritical power plant and emissions reduction by AI approach. *Energy Conversion and Management* **250,** 114913 (2021).

40. Ashraf, W. M. *et al.* Construction of operational data-driven power curve of a generator by industry 4.0 data analytics. *Energies* **14,** 1227 (2021).

41. Ashraf, W. M. & Dua, V. Machine learning based modelling and optimization of post-combustion carbon capture process using MEA supporting carbon neutrality. *Digital Chemical Engineering* **8,** 100115 (2023).

42. Ashraf, W. M. *et al.* Artificial intelligence enabled efficient power generation and emissions reduction underpinning net-zero goal from the coal-based power plants. *Energy Conversion and Management* **268,** 116025 (2022).

43. Krzywanski, J. *et al.* AutoML-based predictive framework for predictive analysis in adsorption cooling and desalination systems. *Energy Science & Engineering* **12,** 1969–1986 (2024).

44. Chou, J.-S. & Bui, D.-K. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy and Buildings* **82,** 437–446 (2014).

45. Ashraf, W. M. & 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 Conversion and Management* **314,** 118645 (2024).

46. Ashraf, W. M. *et al.* Artificial intelligence modeling-based optimization of an industrial-scale steam turbine for moving toward net-zero in the energy sector. *ACS omega* **8,** 21709–21725 (2023).

47. Ashraf, W. M. & Dua, V. Storage of weights and retrieval method (SWARM) approach for neural networks hybridized with conformal prediction to construct the prediction intervals for energy system applications. *International Journal of Data Science and Analytics,* 1–15 (2024).

48. Vogel, G., Balhorn, L. S. & Schweidtmann, A. M. Learning from flowsheets: A generative transformer model for autocompletion of flowsheets. *Computers & Chemical Engineering* **171,** 108162 (2023).

49. Schweidtmann, A. M. Generative artificial intelligence in chemical engineering. *Nature Chemical Engineering* **1,** 193–193 (2024).

50. Jablonka, K. M. *et al.* Machine learning for industrial processes: Forecasting amine emissions from a carbon capture plant. *Science Advances* **9,** eadc9576 (2023).

51. Min, Q. *et al.* Machine learning based digital twin framework for production optimization in petrochemical industry. *International Journal of Information Management* **49,** 502–519 (2019).

52. Ibrahim, M. S., Dong, W. & Yang, Q. Machine learning driven smart electric power systems: Current trends and new perspectives. *Applied Energy* **272,** 115237 (2020).

53. Deon, B. *et al.* Digital twin and machine learning for decision support in thermal power plant with combustion engines. *Knowledge-Based Systems* **253,** 109578 (2022).

54. Boukerche, A. & Wang, J. Machine learning-based traffic prediction models for intelligent transportation systems. *Computer Networks* **181,** 107530 (2020).

55. Hou, Z.-S. & Wang, Z. From model-based control to data-driven control: Survey, classification and perspective. *Information Sciences* **235,** 3–35 (2013).

56. Hewing, L. *et al.* Learning-based model predictive control: Toward safe learning in control. *Annual Review of Control, Robotics, and Autonomous Systems* **3,** 269–296 (2020).

57. Wang, C. & Hill, D. J. Learning from neural control. *IEEE Transactions on Neural Networks* **17,** 130–146 (2006).