Energy and AI 16 (2024) 100363

Contents lists available at ScienceDirect

Energy and AI

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

Data Information integrated Neural Network (DINN) algorithm for modelling and interpretation performance analysis for energy systems

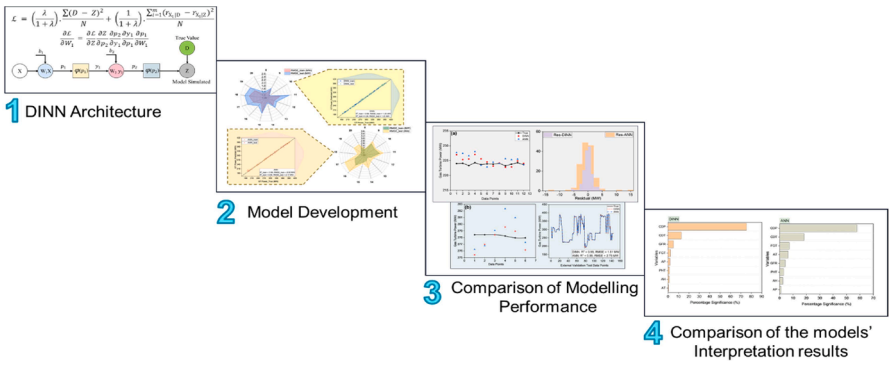
Waqar Muhammad Ashraf *, Vivek Dua *

The Sargent Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK

HIGHLIGHTS

- Data Information integrated Neural Network (DINN) algorithm is proposed.
- Loss function is augmented with the correlation present in the data of the variables.
- DINN offers improved modelling performance than ANN model.
- DINN based interpretations are qualitatively backed by domain-knowledge.
- Integrating the data information can enhance the machine learning's interpretability.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Explainable AI
Model interpretation
Scientific machine learning
Artificial neural network
Loss function

ABSTRACT

Developing a well-predictive machine learning model that also offers improved interpretability is a key challenge to widen the application of artificial intelligence in various application domains. In this work, we present a Data Information integrated Neural Network (DINN) algorithm that incorporates the correlation information present in the dataset for the model development. The predictive performance of DINN is also compared with a standard artificial neural network (ANN) model. The DINN algorithm is applied on two case studies of energy systems namely energy efficiency cooling (ENC) & energy efficiency heating (ENH) of the buildings, and power generation from a 365 MW capacity industrial gas turbine. For ENC, DINN presents lower mean RMSE for testing datasets (RMSE_test = 1.23 %) in comparison with the ANN model (RMSE_test = 1.41 %). Similarly, DINN models have presented better predictive performance to model the output variables of the two case studies. The input perturbation analysis following the Gaussian distribution for noise generation reveals the order of significance of the variables, as made by DINN, can be better explained by the domain knowledge of the power generation operation of the gas turbine. This research work demonstrates the potential advantage to integrate the information present in the data for the well-predictive model development complemented with improved interpretation performance thereby opening avenues for industry-wide inclusion and other potential applications of machine learning.

* Corresponding authors at: The Sargent Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK.
E-mail addresses: waqar.ashraf.21@ucl.ac.uk (W.M. Ashraf), v.dua@ucl.ac.uk (V. Dua).

https://doi.org/10.1016/j.egyai.2024.100363

Available online 25 March 2024

Introduction

The recent advancement in the information communication technologies (ICT) has revolutionized the way the data is generated, stored and made accessible [1]. The higher computational power and improved hardware capabilities enable the widespread exploitation of the available data for the data-driven modelling of the given system in various application domains which has led to the realization of fourth research paradigm often called *Data-Driven Science* we are living, and is followed by *Computational Science* (until 2010), *Theoretical Science* (until 1950) and *Empirical Science* (until 1600). Artificial intelligence-based technologies and machine learning (ML) algorithms remain the key drivers for the advancement in *Data-Driven Science* [2]. Large language models which are classified under the Generative Artificial Intelligence and its recent application called ChatGPT offers potential advantages on how we search the information and consume it for the tasks thereby indicating the usefulness of the language models in the real life [3].

The advanced modelling algorithms of ML can construct the effective functional relationship with the hyper dimensional input space and the predictive accuracy of these models is significantly high compared with those of low-order regression models for nonlinear and complex function space [4]. The predictive mechanism of the ML models is essentially black-box and interpreting how the model predicts is a key challenge in the mass deployment of the ML models in various industrial sectors including chemical process industries, oil and gas and power generation systems. Generally, the physical knowledge is ignored or difficult to incorporate during the ML model development [5] and thus decision makers find the models untrustworthy for their deployment into the industrial environments where a slight mistake can be catastrophic [5].

The scientific community has devised techniques to explain the model's predictions and has classified them into Ante-hoc and Post-hoc approaches [6]. The ML models built under the Ante-hoc class are interpretable by design but have low modelling accuracy [7]. On the other hand, post-hoc interpretable techniques are applied after the black-box model development to understand the significance / contribution of the causal variables towards the prediction of output variables [8]. Since, the black-box models offer improved modelling accuracy compared with self-explanatory and low-order regression models, the scientific community is engaged in integrating the available knowledge during the model development for the improved predictive and interpretable performance of the ML models.

Physics informed neural network (PINN) is trained on the modified loss function that is augmented with the governing equations and physical laws [9]. Thus, during the development of PINN, the parameters update is made after satisfying the applied constraints [10]. However, for many real-life applications, where the first-principle models are not available or the development of the mathematical model is quite difficult given the size and complexity of the considered system, parameters update can be made by integrating the information present in the dataset collected from the considered system. The patterns in the dataset are unique with respect to the state and operating constraints of the system which are sometimes difficult to model by the mathematical equations. Thus, integrating the information present in the data by some relevant statistical measures can be an effective method to guide the parameters update during the ML model development that may exhibit improved interpretability in the post-hoc interpretation analysis.

There are numerous statistical terms that can exhibit the information availability and its nature present in the dataset in different ways. Out of them, Pearson correlation coefficient is a standard term to investigate the linear relationship between the two variables considering the available data associated with the variables. The data-driven information as quantified by the Pearson correlation coefficient represents the behaviour of the system under different modes of operation that is useful to be integrated for building accurate functional map within the ML model. Further, the interactions that exist between the pair of variables are captured by the Pearson correlation coefficient that can support the modelling and interpretation performance of the ML model for the function space that is highly nonlinear and is difficult to be approximated for the applications. In some recent studies carried out on object recognition, image processing and other applications [11–16], the correlation information is exploited for the improved predictive and interpretation performance of the ML models. However, integrating the correlation information during the artificial neural network (ANN) model development for regression-based problems is not reported in the literature. Furthermore, the impact of integrating the data-driven correlation information on the modelling and interpretation performance of the model for the regression-based problems associated with the energy systems is also lacking in the literature that needs to be investigated to contribute to the potential solution for the black box nature of ANN model.

In this work, we present Data Information integrated Neural Network (DINN) algorithm that attempts to integrate the data information in the form of correlation coefficient to enhance the model's predictive and interpretation performance. Considering the potential benefits of integrating the correlation information for the improved modelling performance of the neural networks [13,15,16], it is hypothesized that the available correlation information, that captures the data-driven relationships between the variables, may contribute to enhance the modelling and interpretation performance of the ML modelling algorithms for the energy systems. Inspired by the working mechanism of PINN, the loss function of DINN is customized to integrate the available data information by Pearson correlation coefficient that contributes to the parameters (weights and biases) update mechanism during the model development. Further, a constraint is added on the stopping condition for the DINN training that minimizes the absolute deviation for the correlation computed between the input variable with those of true output variable and model-simulated observations for the output variable. Thus, it is expected that DINN trained over the customized cost function and stopping condition may offer the improved interpretability performance for the regression-based problems taken from the energy systems as investigated by the available post-hoc techniques.

The proposed DINN algorithm is investigated on the two case studies from the energy systems – energy efficiency cooling & energy efficiency heating of buildings performance taken from University of California Irvine open-source dataset [17,18] and the other case study involves the modelling of power generation from a 395 MW capacity gas turbine by the real industrial dataset. The predictive performance of DINN is compared with a standard artificial neural network (ANN) to investigate the modelling efficacy of the DINN. Furthermore, the input perturbation-based approach, following the Gaussian distribution of sample creation, is implemented to qualitatively compare the interpretability results for DINN and ANN. The qualitative interpretation approach incorporates the domain knowledge of the system to validate the model's interpretation results and is potentially applied by the researchers having the deep understanding of the working of their systems [19]. The novel aspect of this research is to incorporate the inclusion of data information in the form of correlation to guide the parameters update during the ML model development that can be helpful to achieve improved modelling and interpretability performance for energy systems from the ML models thereby advancing the research in the domain of machine learning. Note that this work does not propose a new approach or network for interpretability analysis and instead incorporates the data-based interpretation into the design of DINN, which once has been trained is used for interpretability analysis using traditional techniques. The incorporation of data-based interpretability analysis however typically leads to better model performance and hence improved interpretability.

Methods

Mathematical expression for DINN

Let us consider that X is the matrix of input variables, $X = [X_1, X_2,.., X_m]$ having the dimension of m by N where 'N' represents the total number of observations associated with X. The set X is deployed to make the functional map with the output variable D (1 by N dimension) by the DINN model. The number of neurons in the input layer of DINN are specified by the elements in X. W_1 is a matrix having the weight connections from the input to hidden layer (size 'h') of DINN, and W_1 has the dimension of m by h. Whereas, the weight connections from the hidden layer to output layer of DINN are enclosed in W_2 and it has the dimension of 1 by h. The generic information flow and processing taking place along the architecture of DINN is shown graphically on Fig. 1. However, the mathematical computations occurring in DINN are as follows:

$$p_1 = \sum W_1 \odot X^T + b_1 \tag{1}$$

$$y_1 = f_1(p_1) \tag{2}$$

here, b_1 is the matrix of the bias values embedded at the hidden layer neurons of DINN and it has the dimension of m by 1; $W_1 \odot X$ is the elemental multiplication of X with the associated weight connections in W_1 and the summation ($p_1 = \sum W_1 \odot X + b_1$) is computed at the hidden layer neurons; f_1 is the activation function applied on hidden layer and transforms p_1 onto the scale of the activation function. In DINN, we have applied tangent hyperbolic activation function on the hidden layer that scales down p_1 nonlinearly into -1 to 1 and stores it in y_1. The information (y_1) from the hidden to output layer of DINN is transmitted for further processing which is expressed as follows:

$$p_2 = \sum W_2 \odot y_1 + b_2 \tag{3}$$

$$Z = f_2(p_2) \tag{4}$$

Here, b_2 is a bias matrix initiated at the output layer having the dimension of 1 by 1; $W_2 \odot y_1$ is the elemental multiplication between y_1 and W_2 (weight connections from the hidden to output layer of DINN). The summation ($p_2 = \sum W_2 \odot y_1 + b_2$) is calculated at the output layer of DINN and is transformed by the activation function f_2 to produce the model-simulated response (Z) from DINN. In this work, we have applied linear activation function (f_2) on the output layer of DINN.

The loss function ($\mathscr{L}$) constructed in this work includes the mean square of error (true output variable (D) and model-simulated response (Z)), and the mean squared deviation between the Pearson correlation computed for the elements of X_i with respect to D ($r_{X_i|D}$) and Z ($r_{X_i|Z}$). The Pearson correlation coefficient measures the linear dependence between the two variables and integrating it in the loss function introduces an extra term to minimize deviation between the correlation value for the true and model-simulated responses. The significance of adding a correlation coefficient term is two fold: (i) it steers the computation of model parameters so as to improve model predictions of the trained model by extracting and then incorporating the information that exists in the data, and (ii) the user already has some insight into the interdependence of the input-output variables even before the network is trained and that insight is built into model training, resulting in DINN model that provides, after suitable analysis, a better interpretation than the traditional ANN. Thus, the model-simulated responses improve the correlation with X_i, that was initially present in the dataset, during the iterative training of the model. The customized loss function can lead to construct the efficient functional map between the input-output variables for the trained DINN algorithm which in turn may offer the improved interpretation performance in comparison with those of a standard ANN model that lacks the correlation information in its loss function.

The Pearson correlation coefficient varies from +1 (strong positive correlation) to -1 (strong negative correlation). Whereas, zero value of Pearson correlation coefficient indicates the absence of any linear dependence between the two variables. It is important to mention here that the input variables, which are relevant to model the output variable, should be selected carefully considering the domain-knowledge or process expertise so that correlation information available for the pair of variables is relevant and can contribute to development of the well-performing model. The customized loss function for DINN is different from those of a traditional ANN model (ANN) which generally contains a single error term. However, integrating the correlation term in the customized loss function contributes to the parameters update in the iterative training of DINN which can enable the network to possess improved modelling as well as interpretation performance. The two terms in the proposed $\mathscr{L}$ are weighted by λ and are written as follows.

$$\mathscr{L} = \left(\frac{\lambda}{1+\lambda}\right).\frac{\sum (D - Z)^2}{N} + \left(\frac{1}{1+\lambda}\right).\frac{\sum_{i=1}^{m} \left(r_{X_i|D} - r_{X_i|Z}\right)^2}{N} \tag{5}$$

λ is a hyperparameter that decides the contribution of $\frac{\sum (D - Z)^2}{N}$ and $\frac{\sum_{i=1}^{m} (r_{X_i|D} - r_{X_i|Z})^2}{N}$ to compute the loss function. The two error terms in $\mathscr{L}$ are weighted by λ and the user can decide the weighted values for the two error terms based on the value of λ. The two terms augmented in $\mathscr{L}$ tends to minimize the deviation between the true and model-simulated responses and synergizes to minimize $\mathscr{L}$ for building a well-trained DINN model. The gradient descent with momentum algorithm is uti-

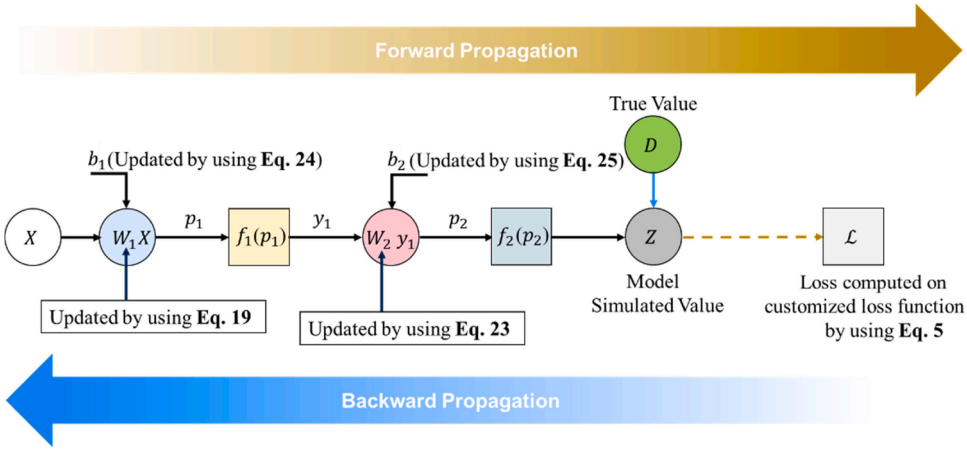


Fig. 1. The generic schematic on the information flow and processing taking place in DINN.

lized to achieve the optimum value of the parameters, i.e., weight and bias since the algorithm offers fast as well as stable error convergence in comparison with the performance of the gradient descent algorithm [20]. The partial derivative of $\mathscr{L}$ with respect to the parameter (W_1, W_2, b_1, b_2) is taken, the information is propagated backward and integrated with the gradient descent with momentum algorithm for the parametric update in an iterative approach to achieve the minimum of $\mathscr{L}$.

The update for weight connections (W_1) by gradient descent with momentum algorithm is given as:

$$W_1^{\text{new}} = W_1 - \eta V_{W_1} \tag{6}$$

Where, η is the learning rate parameter and V_{W_1} is the velocity matrix that is defined as [21]:

$$V_{W_1} = \beta V_{W_1} + (1-\beta)\frac{\partial \mathscr{L}}{\partial W_1} \tag{7}$$

Here, β is the momentum parameter and V_{W_1} is initialized as a zero-matrix having the same dimension as that of W_1. $\frac{\partial \mathscr{L}}{\partial W_1}$ is computed using the chain-rule:

$$\frac{\partial \mathscr{L}}{\partial W_1} = \frac{\partial \mathscr{L}}{\partial Z}\frac{\partial Z}{\partial p_2}\frac{\partial p_2}{\partial y_1}\frac{\partial y_1}{\partial p_1}\frac{\partial p_1}{\partial W_1} \tag{8}$$

$$\frac{\partial \mathscr{L}}{\partial Z} = \frac{\partial}{\partial z}\left(\frac{\sum (D-Z)^2}{N} + \frac{\sum_{i=1}^{m}\sum \left(r_{X_i|D} - r_{X_i|Z}\right)^2}{N}\right) = \frac{-2(D-Z)}{N} - r_{X_i|Z}.\left(\frac{X_i^{\mu}}{\mathscr{B}_i} - \frac{Z^{\mu}}{\mathscr{M}}\right) \tag{9}$$

where,

$$X_i^{\mu} = X_i - \overline{X}_i \tag{10}$$

$$Z^{\mu} = Z - \overline{Z} \tag{11}$$

$$\mathscr{B}_i = \sum_i^m (X_i - \overline{X}_i).(Z - \overline{Z}) \tag{12}$$

$$\mathscr{M} = \sum (Z - \overline{Z})^2 \tag{13}$$

$$\frac{\partial Z}{\partial p_2} = \frac{\partial p_2}{\partial p_2} = 1 \tag{14}$$

$$\frac{\partial p_2}{\partial y_1} = \frac{\partial}{\partial y_1}(W_2 \odot y_1 + b_2) = W_2 \tag{15}$$

$$\frac{\partial y_1}{\partial p_1} = \frac{\partial}{\partial p_1}\left(\frac{e^{p_1} + e^{-p_1}}{e^{p_1} - e^{-p_1}}\right) = 1 - y_1^2 \tag{16}$$

$$\frac{\partial p_1}{\partial W_1} = \frac{\partial}{\partial W_1}(W_1 \odot X + b_1) = X \tag{17}$$

Putting eq. (9-17) in eq. (8):

$$\frac{\partial \mathscr{L}}{\partial W_1} = -\left(\frac{2(D-Z)}{N} + r_{X_i|Z}.\left(\frac{X_i^{\mu}}{\mathscr{B}_i} - \frac{Z^{\mu}}{\mathscr{M}}\right)\right) W_2^T (1 - y_1^2) X^T \tag{18}$$

From eq. (7) and eq. (18), eq. (6) is written as:

$$W_1^{\text{new}} = W_1 + \eta\left(\beta V_{W_1} + (1-\beta)\left(\frac{2(D-Z)}{N} + r_{X_i|Z}.\left(\frac{X_i^{\mu}}{\mathscr{B}_i} - \frac{Z^{\mu}}{\mathscr{M}}\right)\right) W_2^T (1 - y_1^2) X^T;\right) \tag{19}$$

Similarly, the update in the weight connections from hidden to output layer neuron of DINN is written as [21]:

$$W_2^{\text{new}} = W_2 - \eta V_{W_2} \tag{20}$$

$$V_{W_2} = \beta V_{W_2} + (1-\beta)\frac{\partial \mathscr{L}}{\partial W_2} \tag{21}$$

Here, V_{W_2} has the same dimensions as W_2 and is initialized as zero matrix. We can expand $\frac{\partial \mathscr{L}}{\partial W_2}$ by applying the chain rule as follows:

$$\frac{\partial \mathscr{L}}{\partial W_2} = \frac{\partial \mathscr{L}}{\partial Z}\frac{\partial Z}{\partial p_2}\frac{\partial p_2}{\partial W_2} = -\left(\frac{2(D-Z)}{N} + r_{X_i|Z}.\left(\frac{X_i^{\mu}}{\mathscr{B}_i} - \frac{Z^{\mu}}{\mathscr{M}}\right)\right) y_1 \tag{22}$$

From eq. (21) and eq. (22), eq. (20) is written as:

$$W_2^{\text{new}} = W_2 + \eta(\beta V_{W_2} + (1-\beta)\left(\frac{2(D-Z)}{N} + r_{X_i|Z}.\left(\frac{X_i^{\mu}}{\mathscr{B}_i} - \frac{Z^{\mu}}{\mathscr{M}}\right)\right) y_1\Big) \tag{23}$$

Applying the same methodology, the bias update at the hidden (b_1) and output layer (b_2) is expressed as:

$$b_1^{\text{new}} = b_1 + \eta\left(\beta V_{b_1} + (1-\beta)\left(\frac{2(D-Z)}{N} + r_{X_i|Z}.\left(\frac{X_i^{\mu}}{\mathscr{B}_i} - \frac{Z^{\mu}}{\mathscr{M}}\right) W_2^T (1 - y_1^2)\right)\right) \tag{24}$$

$$b_2^{\text{new}} = b_2 + \eta\left(\beta V_{b_2} + (1-\beta)\left(\frac{2(D-Z)}{N} + r_{X_i|Z}.\left(\frac{X_i^{\mu}}{\mathscr{B}_i} - \frac{Z^{\mu}}{\mathscr{M}}\right)\right)\right) \tag{25}$$

The parameters update is continued such that one of stopping conditions is achieved, i.e., slope is less than 0.00000001, loss value on the testing dataset is less than goal value and the absolute deviation between $r_{X_i|D}$ and $r_{X_i|Z}$, i.e., $|r_{X_i|D}| - |r_{X_i|Z}|$ is minimized to zero. Minimizing the absolute deviation between $r_{X_i|D}$ and $r_{X_i|Z}$ steers the DINN algorithm to approximate the distribution profile of true observations of the output variable against the set of input variables that enhances the accuracy of the functional mapping between the input and output variables. The stopping constraint, i.e., $|r_{X_i|D}| - |r_{X_i|Z}|$ can support the improved modelling and interpretability performance of the DINN model since the correlation information available from the dataset is explicitly exploited for the model development in comparison to ANN which does not have additional constraint to be satisfied prior to the training of model is finished.

Evaluation criteria

Two statistical measures, namely coefficient of determination (R^2) and root-mean-squared-error (*RMSE*) are utilized to investigate the predictive performance of the DINN. The mathematical expression for R^2 and *RMSE* are provided as follows:

$$R^2 = 1 - \frac{\sum_i^N (Z_i - D_i)^2}{\sum_i^N (D_i - \overline{D}_i)^2} \tag{26}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Z_i - D_i)^2} \tag{27}$$

here, Z_i and D_i are the model-simulated responses and true value of output variables respectively for $i = 1,2,3,\ldots N$ equal to total number of observations. R^2 measures the predictive accuracy of the model and it varies from zero (poor predictive performance of the model) to one (perfect match between the true and model-predictive responses). On the other hand, *RMSE* indicates the mean deviation between the true and model-predicted responses and is made as low as possible to achieve the good functional map between the input and output variable of the DINN.

Results and discussion

Case studies from energy systems

The algorithm of DINN is applied on two case studies – one is taken

from a benchmark dataset on energy efficiency cooling (ENC) & energy efficiency heating (ENH), and the second case study involves modelling the power generation from an industrial-scale 395 MW capacity gas turbine. Thus, DINN has been implemented on the example datasets that are either benchmarked in literature and the modelling performance of the DINN is also investigated on industrial application as well. It is also important to mention here that modelling performance of DINN is compared with those of an ANN model which is trained on a loss function comprising on mean-squared-error terms. It allows to investigate the effect of modification of the loss function on the modelling performance of the DINN for the case studies. Both DINN and ANN models are trained in MATLAB 2019 b software. The comparative performance results for DINN and ANN are provided in the following.

Case study – 1: energy efficiency cooling & energy efficiency heating

The energy efficiency cooling (ENC) and energy efficiency heating (ENH) datasets for the buildings consists of eight input variables namely relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area and glazing area distribution. The dataset has 768 observations associated with the variables. 80 % of data is deployed for training while remaining 20 % data is used for testing purpose during the models development for ENC and ENH. The number of neurons in the hidden layer are kept at 10 to compare the modelling performance of the DINN and ANN with those of feed forward neural network (FFNN) reported in literature [22] and are trained 20 times to average out the performance of the models considering the different parameters initialization. Whereas, different values of λ are explored in the range of 0.1 to 0.2 for the training of DINN.

Fig. 2 compares the modelling performance of DINN and ANN models to model ENC and ENH. The performance metrics are measured on the training and testing datasets for the two output variables. The range in variation in the RMSE computed for DINN and ANN on the test dataset for ENC and ENH are graphically visualized by box plot on Fig. 2 (a). It is noted that mean RMSE of DINN on test dataset for ENC and ENH are 1.23 % and 0.56 % which are comparatively lower than those of ANN, i.e., 1.41 % and 0.65 % respectively. Furthermore, out of the trained ANN and DINN models, the performance metrics computed on true and model predicted responses for comparatively better model of ENC and ENH are presented on Fig. 2(b). In case of ENC, R^2 remains quite comparable both on training and testing datasets for DINN and ANN models. However, DINN based RMSE for training and testing dataset are 1.04 % and 1.14 % respectively which are lower than those of ANN model, i.e., RMSE_train = 1.08 % and RMSE_test = 1.24 %. The similar observation can be noted for the modelling performance of DINN and ANN for ENH where R^2 for training dataset is comparable for the models. However, DINN based predictions on training and testing dataset present lower RMSE in comparison with those of ANN. The modelling performance of the two models for ENH is summarized as follows: $(\text{RMSE\_train} = 0.46\ \%)_{\text{DINN}} < (\text{RMSE\_train} = 0.51\ \%)_{\text{ANN}}$ and $(\text{RMSE\_test} = 0.43\ \%)_{\text{DINN}} < (\text{RMSE\_test} = 0.52\ \%)_{\text{ANN}}$. Furthermore, modelling performance of DINN are also compared with those of FFNN for ENC and ENH as reported in literature [22]. The authors trained the FFNN 20 times with the same architectural configuration of the model and reported the mean results for the repetitive training of models as follows: RMSE_test = 1.63 % and RMSE_test = 0.63 % for ENC and ENH respectively [22] which are comparatively higher than those of DINN as described above. Comparing the predictive performance of DINN with those of ANN and FFNN, it is found that DINN has superior modelling performance for the ENC and ENH that can be attributed to integrating the data information in the form of correlation between the input and output variable that might have helped to optimize the parameters effectively to achieve the better predictive performance of DINN.

The input variables for ENC and ENH have inter-dependencies as

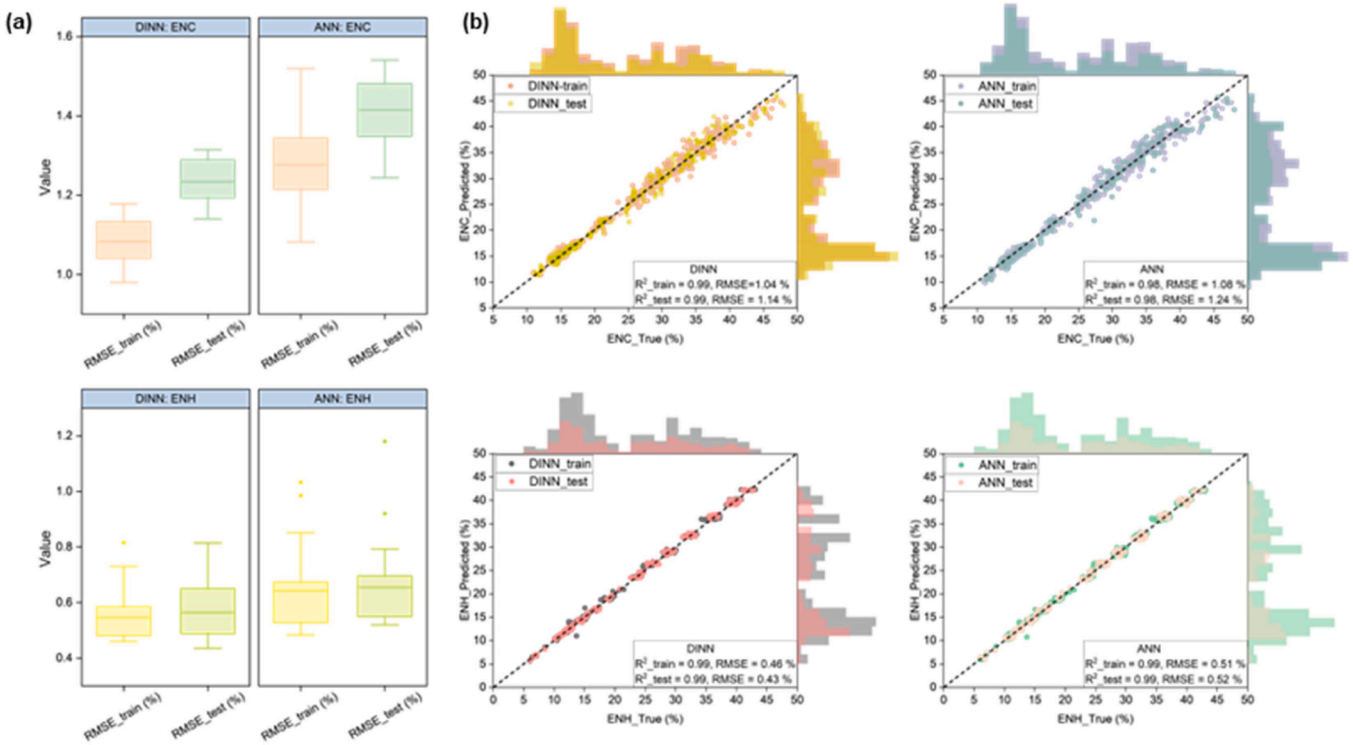


Fig. 2. Comparison of modelling performance of the DINN with the ANN model for ENC, and b) ENH. The graphical visualization of variability in the RMSE on the test dataset computed for 20 times training of ANN and DINN models under same initial conditions for ENC and ENH is presented. (b) True vs predicted responses of DINN and ANN having the comparatively better performance metrics out of the 20 trained models. The performance metrics computed on testing dataset for DINN are appeared to be better than those of ANN for ENC and ENH thereby showing improved modelling performance of DINN. The bars along the edges of the graph represent the data-distribution density profiles for the training and testing dataset with respect to true and model-predicted responses.

represented in the dataset in the form of the operating levels of the variables. However, the two input variables namely orientation and glazing area have relatively fewer operating levels and interactions with the other input variables in the test dataset. This allows to study the variation in the two said input variables considering the test dataset of DINN and ANN on the ENC and ENH and compares the model predicted responses with the true responses. This parametric study explores the local regions of the test dataset to investigate the predictive performance of the models as well as the model's ability to generalize the relationships with the variables. DINN and ANN based responses for ENC and ENH on the variation in orientation and glazing area are presented on Fig. 3. DINN and ANN based responses are depicted by red and blue lines respectively while the true responses are shown by solid black lines. The responses as made by the DINN and ANN with respect to the variation in the two variables are closely related with the true observations that indicate the good modelling performance of the two models for the testing dataset. However, closely comparing the predictive performance of the two models, it is apparent that DINN has improved prediction profiles than those of the ANN. This demonstrates the better predictive and generalization ability of the DINN compared with the ANN model and depicts the competitive advantage to integrate the correlation information in the training of DINN model for the improved predictive analytics. Furthermore, an improved predictive performance of the DINN on the parametric variation in the two variables also contributes to understand the model's interpretability and evaluating the accuracy of the model's predictions by comparing with the true observations.

Case study – 2: power generation from an industrial gas turbine

The power production from an industrial-scale gas turbine is maintained under the operation management of a number of sub-systems integrated with the power generation operation. The variables have nonlinear and complex interactions and thus affect the power generation operation. In this work, we have selected the input variables based on the domain knowledge of the power plant operation and literature review [23,24]. Therefore, compressor discharge air temperature (CDT), compressor discharge air pressure (CDP), fuel gas flow rate (GFR), performance heater gas outlet temperature (PHT), fuel gas temperature (FGT), ambient temperature (AT), ambient pressure (AP) and ambient humidity (AH) are deployed to construct the data-driven process model for the power generation.

The dataset containing 578 observations is taken from the gas power plant and the data split ratio of 0.8 and 0.2 is applied for training and testing purpose respectively. The number of neurons in the hidden layer is a critical hyperparameter that controls the complexity embedded in the processing elements in the ANN to approximate the given function with reasonable accuracy. Thus, we have explored the design space for the hidden layer neurons varying from 1 × to 2.5 × times of the neurons in the input layer [25]. Whereas, λ is kept at 0.4 for training the DINN. The performance metrics are measured in the training and testing phase of the model development for DINN. Fig. 4 shows the modelling performance of the DINN and ANN with respect to the effect of number of hidden layer neurons on the RMSE of the two models as presented on the spider web plots. Closely comparing the predictive performance of the models in the training and testing phase, it is apparent that DINN having 10 hidden layer neurons achieves the lowest RMSE both for training and testing datasets in comparison of other DINN models having different number of hidden layer neurons. Similarly, the lowest RMSE is observed corresponding to 18 hidden layer neurons in ANN model. The mapping of true and predicted values of gas turbine power (GT Power) for optimal hidden layer configuration for DINN and ANN are presented on Fig. 4. Both DINN and ANN has comparable value of R^2 in the training and testing datasets. However, the predictive performance of the two models differs significantly when measured by RMSE for training and testing datasets. For DINN, RMSE_train = 1.20 MW and RMSE_test = 1.25 MW is observed which is lower than those of the ANN model, i.e., RMSE_train = 2.0 MW and RMSE_test = 2.17 MW. The performance comparison of the DINN and ANN model reveals that DINN presents better predictive performance, both for training and testing datasets to model the power generation from a gas turbine.

In the previous section, we have presented the modelling performance of the DINN and ANN model during their development and the performance was measured both for training and testing datasets. Recently, researchers have also investigated the validation of the machine learning models by deploying them to predict the external validation dataset – potentially unseen dataset taken from the system under consideration. Thus, external validation test offers a rigorous testing step to evaluate the generalization of the trained machine learning models [26,27]. Therefore, the external validation dataset consisting of 144 randomly selected observations from the power generation operation of the gas turbine are taken from the power plant and are deployed to be predicted from the trained DINN and ANN models. The predicted responses from the two models are compared with the true observations, performance metrics, i.e., R^2 and RMSE are calculated, and the residuals are shown on Fig. 5(a). Comparing the performance metrics for the DINN and ANN, it is noted that the two models have comparable values of R^2 for external validation dataset. However, RMSE measured for DINN is lower than that of ANN model (RMSE_DINN = 1.51 MW < RMSE_ANN = 2.75 MW). Furthermore, 92.4 % and 76.4 % of the residuals are distributed from -2.5 MW to 2.5 MW range for DINN and

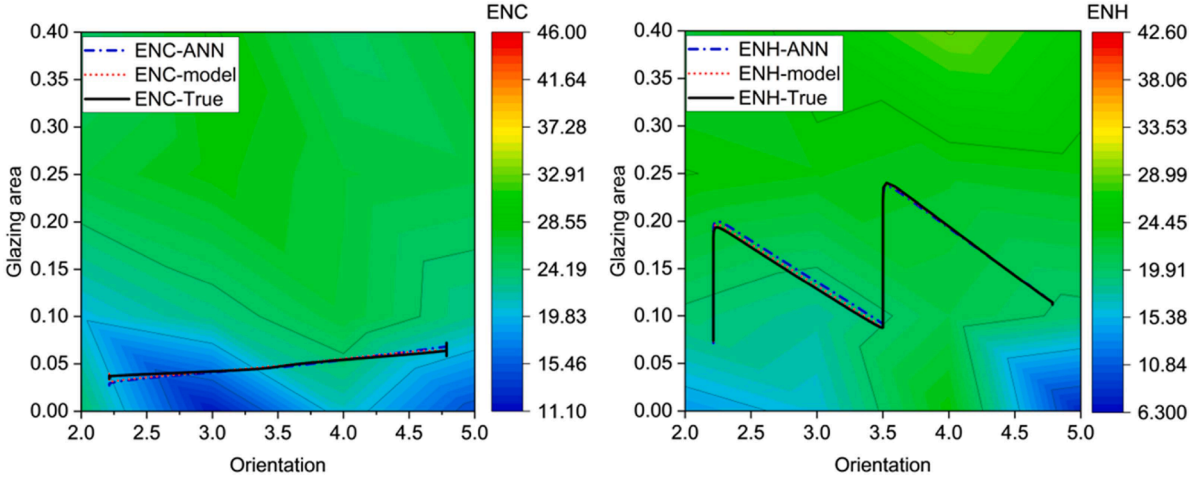


Fig. 3. Effect of varying orientation and glazing area on the ENC and ENH performance as studied by DINN and ANN model. The model predicted responses are compared with the true values to evaluate the predictive accuracy of the models.

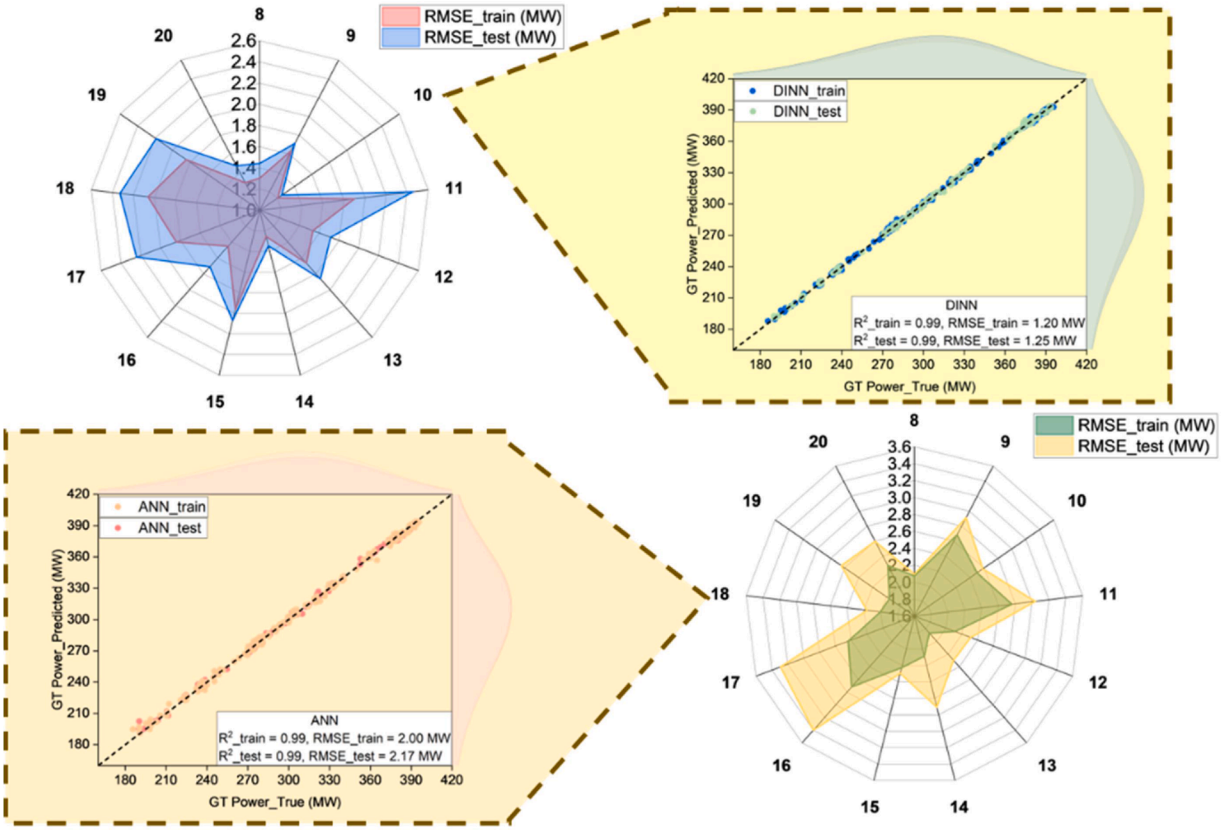


Fig. 4. Predictive performance of DINN and ANN to model the power generation from an industrial-scale 395 MW capacity gas turbine. The performance metrics are measured for training and testing datasets to investigate the effectiveness of the functional map created between the input and output variables for DINN and ANN. The data-distribution density profiles for the training and testing dataset with respect to true and model-predicted responses are drawn along the edges of the graph.

ANN model respectively which depicts that DINN has relatively lower magnitude of the residuals in comparison with those of ANN for the reasonably-wide residual range.

The predictive performance of the two models is also evaluated in the local space of gas turbine power considering the variation in ambient temperature and ambient humidity such that the other input variables are nearly constant on their operating values. Since, the ambient conditions are changed independent of the other input variables during the power generation, therefore the impact of the ambient temperature and ambient humidity on the state of power generation is available in the dataset that serves as a ground truth. This allows us to investigate the predictive performance of the trained DINN and ANN model in the local regions of the output variable and compare it with the true values to evaluate the predictive performance of the trained models. Fig. 5(b) presents the mapping of the DINN and ANN based responses with the true power generation value under two different power generation mode of the gas turbine for a number of input conditions taken from the external validation dataset when ambient temperature and ambient humidity are changed and the remaining input variables are nearly constant. Comparing the distribution of the DINN and ANN model-based responses around the true gas turbine power under two power generation modes, it is noted that in general, DINN seems to have better prediction profiles that lie close to the true observations. In some instances, ANN offers better predictive responses than those of DINN. However, overall DINN based predictive responses are appeared to be relatively closer to the true observations. Thus, the rigorous evaluation of the DINN is made on the external validation test and the predictive performance is compared with those of ANN model to investigate the modelling accuracy of the DINN to predict the power generation from a gas turbine.

Finally, it is important to investigate the variable significance towards the prediction of the output variable considering the black-box nature of the ANN model. The order of the significance of the input variable presents the relative importance of the variable towards the prediction. The strength of the functional map between the input and output variable which is built by the data-driven ML model and investigated by the variable significance analysis, thus the model based predictions can be interpreted [28]. Furthermore, the order of significance of the variables can be explained by the domain knowledge to confirm the correct interpretation of the model [29].

In this work, we have applied input perturbation method to analyse the variable significance and have utilized the qualitative approach to investigate the interpretation performance of the models [19]. The qualitative approach for the model's interpretability analysis integrates the end-user intuitiveness and domain-knowledge to validate the model's interpretation results. In input perturbation method, the variable, whose significance is to be evaluated, undergoes parametric variation from its minimum to maximum operating level whereas the other input variables are kept at their mean value [30]. Other schemes for the construction of experiments can be followed for the input perturbation method. In the next step, input perturbation method introduces the noise for the given input condition and averages out the responses associated with the output variable to account for the variation in the operating level of the input variables. This offers an extensive investigation at each operating level for the number of noise observations to find the significance of the variable. Thus, the process is repeated for the complete parametric variation in the variable and applied on all input variables to establish the normalized order of the significance. In this work, we have generated 10,000 observations of noise following the Gaussian distribution with respect to the one percent minimum value of the input variables and are deployed for input perturbation based variable significance analysis to understand the model's interpretability for

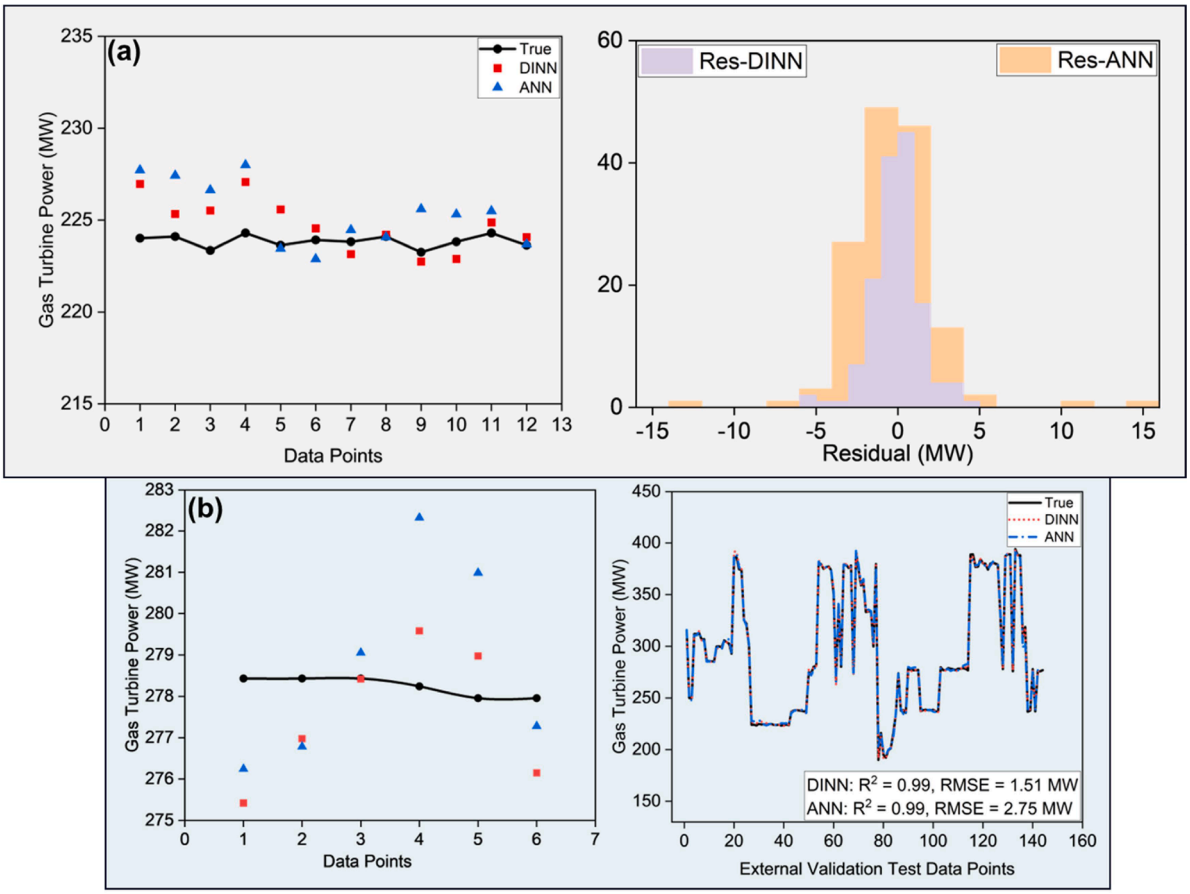


Fig. 5. Predictive performance of DINN and ANN for the external validation dataset. (a) DINN and ANN model-based responses mapped against the true observation for the external validation dataset along with the residuals distributions are shown. (b) The predictive performance of DINN and ANN in the local space of gas turbine power built with the significant contribution of ambient temperature and ambient humidity is presented. The models-predicted responses are plotted against the true observations of the gas turbine power to show the predictive accuracy of the two models.

the predictions. The Gaussian distribution for the noise observations is considered since the sensor measurements generally follow the gaussian distribution for the variables associated with the power generation from the gas turbine.

Fig. 6 shows the order of significance of the input variables on the power generation by input perturbation method as carried out by using DINN and ANN model. The percentage significance of the input variables is computed by the two models for the comparative analysis. CDP and CDT are turned out to be the two most significant input variables affecting the power generation from the gas turbine and their percentage significance is 75.4 % & 12.5 % and 57.7 % & 18.3 % respectively. However, the third significant variable for DINN is GFR having

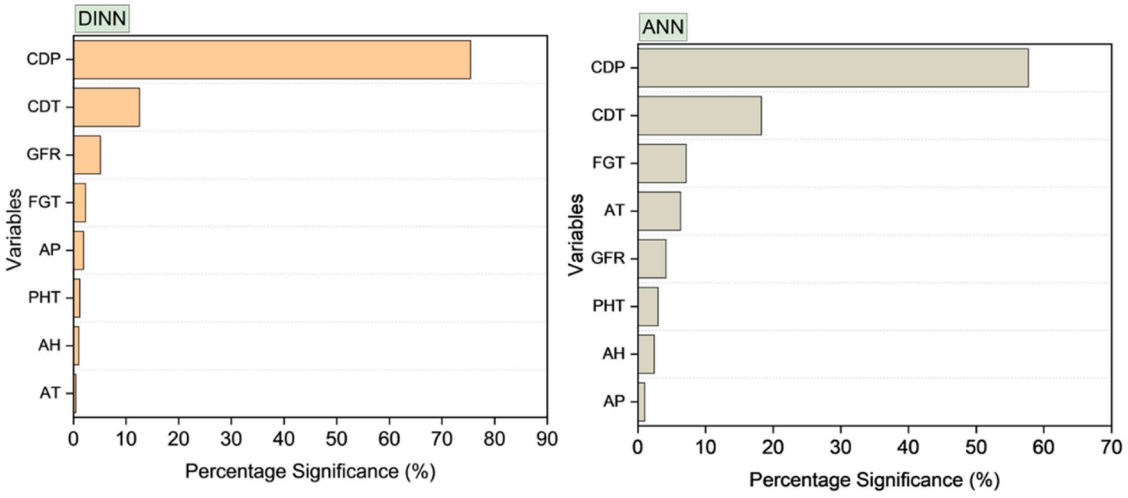


Fig. 6. The significance of the input variables to predict the power generation from the gas turbine. The variables significance analysis presents the importance of the variable towards the prediction of the response variable thereby contributes to the model interpretation.

percentage significance of 5.1 %. On the other hand, FGT is the third significant variable for ANN model with percentage significance value of 7.1 %. The first three input variables, namely CDP, CDT and GFR contributes around a total of 95 % significance towards the prediction of gas turbine power by DINN. Whereas, CDP, CDT, FGT, AT and GFR adds up to approximately 95 % significance towards the prediction of gas turbine power. CDP and CDT are the variables that indicate the conditions of the air at the discharge of compressor. Similarly, GFR represents the gas flow rate entering the combustion chamber.

The fuel combustion in the presence of air produces high-temperature flue gas that expands in the gas turbine to produce power. Thus, the power generation is sensitive to the operating conditions of air and fuel that produces the flue gas for the power production as it is an established working principle for the power generation operation from the gas power plant. Thus, third order of GFR, as established by DINN, is well aligned with the operational and domain knowledge of the power generation from the gas turbine in terms of variable's significance to predict the power generation and thus DINN has improved interpretation performance to predict the power generation with reference to variable's significance order. On the other hand, GFR occupies the fifth place in the order of significance, whereas FGT is termed out to be third significant variable for ANN based interpretation analysis which is not well aligned with the domain-knowledge and expertise for the power generation from the gas power plant. The input perturbation approach for the exploration of model's interpretability indicates the advantage of integrating the correlation information for the parameters update during the DINN development that is represented in the form of improved interpretability towards the model-based predictions. The industrial community and practitioners can benefit from the improved interpretability of the DINN to better formulate the operating strategies for the effective operation management of their industrial complexes, informed decision making and making the simulators for the training of operators and entry-level engineers to better understand the behaviour of system under different operating conditions. Thus, deploying the data information into the modelling algorithm can be helpful to achieve better predictive performance and domain-knowledge backed model's interpretations that contributes to Industry-4.0 vision for smart operation of industrial systems.

Conclusion

The information present in the data can be helpful to achieve the better predictive performance and improved interpretability of the machine learning models. In this paper, we propose a modelling algorithm that integrates the correlation information present in the dataset of the variables and deploys it to train an effective Data Information integrated Neural Network (DINN). The loss function of DINN is augmented with mean squared deviation between the Pearson correlation computed for true and model-simulated responses with the input variables. The parameters update is thus guided by the modified loss function through the gradient descent with momentum algorithm.

The proposed DINN algorithm is applied on two case studies considered from energy systems namely energy efficiency cooling (ENC) and heating (ENH) of buildings, and power generation from a 395 MW capacity gas turbine. The modelling performance of DINN is compared with a standard ANN model. For ENC, DINN presents lower mean RMSE for testing datasets (RMSE_test = 1.23 %) in comparison with the ANN model (RMSE_test = 1.41 %). Furthermore, DINN based predictive performance is found to be lower than those of literature reported results for ENC ($(\text{RMSE\_test} = 1.23\ \%)_{\text{DINN}} < (\text{RMSE\_train} = 1.63\ \%)_{\text{FFNN}}$). Similar results indicating the better predictive performance of the DINN are observed for the case studies.

The external validation test is carried out on the DINN and ANN models for the gas turbine power to investigate the generalization ability of the models. DINN presents the lower prediction error ((RMSE_DINN = 1.51 MW < RMSE_ANN = 2.75 MW) and residuals distribution on a tight residual range than those of ANN. Furthermore, input perturbation-based variables significance analysis is carried out to establish the order of significance for the input variables affecting the power generation from the gas turbine. DINN based variables significance order can be better explained by the domain knowledge of the power generation from the gas turbine.

This research presents the benefits of augmenting the loss function to guide the parameters update for the development of a well-predictive DINN model that also presents improved interpretability. Thus, integrating the data information in the working algorithm of the ML models can potentially improve the model's predictive and interpretation performance.

Data availability

The code developed in this research to perform different tasks can be provided on request.

Funding

Punjab Education Endowment Fund (PEEF) provided funding to Waqar Muhammad Ashraf for pursuing his PhD at University College London.

CRediT authorship contribution statement

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

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Kang C, et al. Big data analytics in China's electric power industry: modern information, communication technologies, and millions of smart meters. IEEE Power and Energy Magazine 2018;16(3):54–65.
[2] Tsihrintzis GA, Sotiropoulos DN, Jain LC. Machine learning paradigms: advances in data analytics. Springer; 2019.
[3] Hopkins AM, et al. Artificial intelligence chatbots will revolutionize how cancer patients access information: chatGPT represents a paradigm-shift. JNCI Cancer Spectr 2023;7(2).
[4] Muhammad Ashraf W, et al. Optimization of a 660 MWe supercritical power plant performance—A case of Industry 4.0 in the data-driven operational management. Part 2. Power generation. Energies (Basel) 2020;13(21):5619.
[5] Chen Z, et al. Interpretable machine learning for building energy management: a state-of-the-art review. Adv Appl Energy 2023:100123.
[6] Antoniadi AM, et al. Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. Appl Sci 2021;11(11):5088.
[7] Lisboa P, et al. The coming of age of interpretable and explainable machine learning models. Neurocomput 2023;535:25–39.
[8] Kenny EM, Keane MT. Explaining Deep Learning using examples: optimal feature weighting methods for twin systems using post-hoc, explanation-by-example in XAI. Knowl Based Syst 2021;233:107530.
[9] Cai S, et al. Physics-informed neural networks (PINNs) for fluid mechanics: a review. Acta Mechanica Sinica 2021;37(12):1727–38.
[10] Almajid MM, Abu-Al-Saud MO. Prediction of porous media fluid flow using physics informed neural networks. J Petroleum Sci Eng, 2022;208:109205.
[11] Li S, et al. Imaging through glass diffusers using densely connected convolutional networks. Optica 2018;5(7):803–13.

[12] Ou X, et al. A Hyperspectral Image Change Detection Framework With Self-Supervised Contrastive Learning Pretrained Model. IEEE J Sel Top Appl Earth Obs Remote Sens 2022;15:7724–40.
[13] Fan W, et al. 2D shape reconstruction of irregular particles with deep learning based on interferometric particle imaging. Appl Opt 2022;61(32):9595–602.
[14] Nagaraju TV, et al. Predicting California Bearing Ratio of Lateritic Soils Using Hybrid Machine Learning Technique. Buildings 2023;13(1):255.
[15] Liu M, et al. BIT-MI Deep Learning-based Model to Non-intrusive Speech Quality Assessment Challenge in Online Conferencing Applications. In: Proc. Interspeech 2022; 2022. p. 3288–92.
[16] Sun X, et al. Dual-task convolutional neural network based on the combination of the U-Net and a diffraction propagation model for phase hologram design with suppressed speckle noise. Opt Express 2022;30(2):2646–58.
[17] Tsanas A, Xifara A. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy Build 2012; 49:560–7.
[18] Oh SJ, Schiele B, Fritz M, et al. Towards Reverse-Engineering Black-Box Neural Networks. In: Samek W, et al., editors. Explainable AI: interpreting, explaining and visualizing deep learning. Editors. Cham: Springer International Publishing; 2019. p. 121–44.
[19] Saleem R, et al. Explaining deep neural networks: a survey on the global interpretation methods. Neurocomput 2022;513:165–80.
[20] Yu H, Wilamowski BM. Levenberg–marquardt training, in intelligent systems. CRC Press; 2018. p. 12-1-12-16.
[21] Ng A. Improving deep neural networks: hyperparameter tuning, regularization and optimization. Deeplearning. ai on Coursera; 2017.
[22] Arnaldo I, O'Reilly UM, Veeramachaneni K. Building predictive models via feature synthesis. In: Proceedings of the 2015 annual conference on genetic and evolutionary computation; 2015.
[23] Qu Z, et al. Prediction of electricity generation from a combined cycle power plant based on a stacking ensemble and its hyperparameter optimization with a grid-search method. Energy 2021;227:120309.
[24] Hundi P, Shahsavari R. Comparative studies among machine learning models for performance estimation and health monitoring of thermal power plants. Appl Energy 2020;265:114775.
[25] Ashraf WM, Dua V. Machine learning based modelling and optimization of post-combustion carbon capture process using MEA supporting carbon neutrality. Digital Chem Eng 2023;8:100115.
[26] Ashraf WM, Dua V. Artificial intelligence driven smart operation of large industrial complexes supporting the net-zero goal: coal power plants. Digital Chem Eng 2023; 8:100119.
[27] Zhang W, et al. Machine learning based prediction and experimental validation of arsenite and arsenate sorption on biochars. Sci The Total Environ 2023;904: 166678.
[28] Kumar R, Singh AK. Chemical hardness-driven interpretable machine learning approach for rapid search of photocatalysts. NPJ Comput Mater 2021;7(1):197.
[29] Zhao S, et al. Interpretable machine learning for predicting and evaluating hydrogen production via supercritical water gasification of biomass. J Clean Prod 2021;316:128244.
[30] Nourani V, Fard MS. Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologic regimes. Adv in Eng Software 2012;47(1):127–46.