Modelling of Engineering Systems with Small Data, a Comparative Study

Morteza Mohammadzaheri*, Payam Soltani
Department of Engineering, Birmingham City University, UK
ORCID IDs (respectively): 0000-0002-8187-6375, 0000-0002-8186-2678
morteza.mohammadzaheri@bcu.ac.uk

Hamidreza Ziaiefar
Department of Process, Energy and Environmental Technology, Porsgrunn, Universitetet i Sørøst-Norge, Norway
ORCID ID: 0000-0001-9137-3176

Mojtaba Ghodsi
School of Energy and Electronic Engineering, University of Portsmouth, Portsmouth, UK
ORCID ID: 0000-0002-8762-9498

Issam Bahadur, Musaab Zarog, Mohammadreza Emadi, Amirhosein Amouzadeh
Mechanical and Industrial Engineering Department, Sultan Qaboos University, Muscat, Oman
ORCID IDs (respectively): 0000-0003-2373-6590,
0000-0002-7325-2634, 0000-0002-1127-1311, 0000-0002-1535-8684

ABSTRACT

This chapter equitably compares five different Artificial Intelligence (AI) techniques for data-driven modelling. All these techniques were used to solve two real-world engineering data-driven modelling problems with small number of experimental data samples, one with sparse and one with dense data. The models of both problems are shown to be highly nonlinear. In the problem with available dense data, Multi-Layer Perceptron (MLP) evidently outperforms other AI models and challenges the claims in the literature about superiority of Fully Connected Cascade (FCC). However, the results of the problem with sparse data shows superiority of FCC, closely followed by MLP and neuro-fuzzy network.

Keywords— Modelling, Artificial Intelligence, Small Data, Sparse Data, Dense Data, Piezoelectric Actuator, Electrical Submersible Pump, ANFIS, FCC Network, MLP, RBFN.

INTRODUCTION

Nowadays, engineering world witnesses two partly conflicting realities:

1. Model-based design/optimisation/control are on rise (Madni & Sievers, 2018; Pal et al., 2022), and analytical and numerical models of many engineering systems cannot serve their purpose satisfactorily, e.g. as detailed in (X. Li et al., 2022; Mohammadzaheri, Grainger, et al., 2012a; Mohammadzaheri, Tafreshi, et al., 2020; Rahbar et al., 2022). This leads to an ascending demand for data-driven models, developed with experimental data.
2. Experiments take time and cost. Hence, experimental data sets often consist of limited number of samples, or they are small.

That is, engineers are more likely to deal with small data rather than big data (Morteza Mohammadzaheri et al., 2018a). Thus, developing accurate models out of small data is a crucial task for engineers (Chang et al., 2015; Zhang et al., 2022). There are a lot of piecemeal research in the literature reporting development of data-driven models for engineering systems with small data, e.g. (Kokol et al., 2022; Liu et al., 2023; Taajobian et al., 2018). However, no comparative research was found on modelling of engineering systems with small data, though a few were found in other areas (Collins et al., 2017; Steyerberg et al., 2000), as the value of models developed with small data is not limited to engineering (Goel et al., 2023; Kitchin & Lauriault, 2015).

AI provides powerful tools to model intricate engineering systems with their input-output data (Castro, 2018; Garg et al., 2015; W. Li et al., 2022). The research question is which AI method suits best to data-driven modelling of engineering systems, when only a small set of data is available. The answer to this question, which necessitate an unprecedented even-handedly comparison of the AI techniques for data-driven modelling of engineering system with small data, is the main contribution of this chapter. In order to answer the aforementioned research question, several AI-based data-driven models were developed with small data to solve two real-world engineering problems: (i) head estimation of an electrical submersible pump (ESP) lifting two-phase petroleum fluids, detailed in (Mohammadzaheri & Ghodsi, 2018) and (ii) selection of the sensing resistor in a charge estimator of a piezoelectric actuator, detailed in (Mohammadzaheri, Emadi, et al., 2019). Neuro-fuzzy and FCC networks, MLPs, and exact and efficient Radial Basis Function Network (RBFN) models as well as linear models have been developed to tackle problems (i) and (ii).

PROBLEM STATEMENT

This section briefly explains dual engineering problems, mentioned in the introduction, which were solved in this research using AI data-driven modelling techniques. The AI techniques will be compared based on their performance in solving these problems:

Head Estimation of Two-Phase Petroleum Fluids Lifted by ESPs

A variety of empirical models are used to estimate head of two-phase petroleum fluids, H_m, lifted by ESPs. Most of them have three inputs. One input is either intake pressure, p_{in}, or density and the other two are among oil flow rate, gas flow rate, mixed fluid flow rate, $q_{m,}$, or gas void ratio, α (volumetric ratio of gas to mixed fluid) (Mohammadzaheri & Ghodsi, 2018). Temperature has been overlooked in this data-driven modelling problem so far (Mohammadzaheri & Ghodsi, 2018). Pump rotational speed definitely affects the head;

however, its role is considered through the affinity law, and the empirical models are usually developed for a single rotational speed (Mohammadzaheri et al., 2016). Inspired by prevalent empirical models, following general model was employed in this work:

$$H_m = f_{ESP}(\alpha, p_{in}, q_{m,}). \quad (1)$$

In this research, the two-phase fluid is a mixture of carbon dioxide and diesel fuel pumped by eight stages of an I-42B radial ESP, as detailed in (Lea & Bearden, 1982; Mohammadzaheri et al., 2015). In total, the results of 109 experiments are available. 74, 17 and 18 data samples were used as modelling, validation and test data sets to identify/approximate and cross-validate f_{ESP}. The exact use of these triple data sets will be detailed in Modelling section. Input space of these data, depicted in Fig.1, is fairly sparse. For instance, few data samples are available from operating areas with high pressure and low flow rate and in operating areas with high gas void ratio. The units of p_{in} and q_m in this research are ksi and gpm (gallons per minute), and α is unit-less.

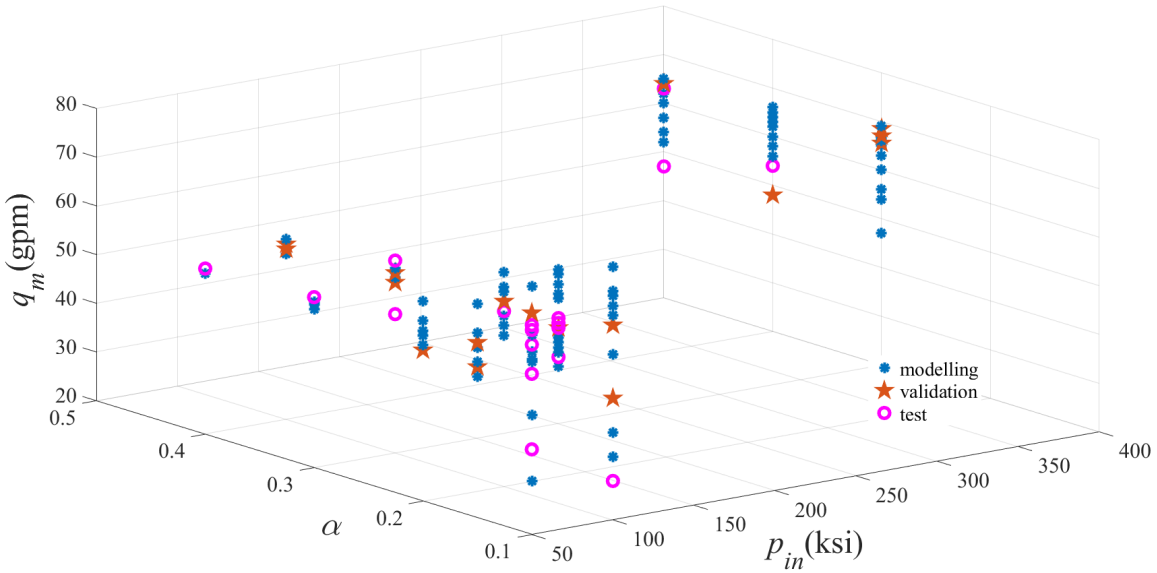

Figure 1. Distribution of modelling, validation and test data to approximate/identify and cross-validate fESP, only input

Selection of Sensing Resistor in a Charge Estimator of a Piezoelectric Actuator

Figure 2 depicts a resistor-based, or digital (Bazghaleh et al., 2013; Mohammadzaheri, Ziaiefar, Ghodsi, et al., 2022), charge estimator of a piezoelectric actuator. The excitation voltage, V_e, is applied on the actuator, leading to a voltage across the sensing resistor, V_S. Since the current passing analogue to digital convertor, A/D, is negligible, the current passing the actuator almost equals the current passing the resistor $=V_S / R_S$. Charge of the actuator is integral of its current. High pass filter removes drift phenomenon as detailed in (Bazghaleh et al., 2018).

According to (Mohammadzaheri, Ziaiefar, & Ghodsi, 2022) as a selection criterion, an apt R_S should lead to a V_S just within the smallest range of A/D input voltage. For instance, if A/D has input voltage ranges of [-1 1] V, [-5 5] V and [-10 10] V, R_S should be selected so that V_S takes the widest possible span within the range of [-1 1] V. However, experiments have shown that a fixed sensing resistor cannot meet the aforesaid criterion for all operating conditions (Mohammadzaheri, Ziaiefar, et al., 2019). In other words, the apt sensing resistor should be selected based on operating conditions e.g. waveform, amplitude and frequency of excitation

voltage (V_e in Fig.2). It has been also shown that analytical models are inaccurate in finding such an apt R_S (Mohammadzaheri, Emadi, et al., 2019). Thus, the remaining alternative is to develop data-driven models to estimate apt R_S based on operating conditions. In this research, it is assumed that V_e is a sinusoidal function of time, with the range of υ (in V) and the frequency of f (in Hz). Hence, υ and f are only operating conditions. That is, (2) is the data-driven model to estimate apt R_S:

$$R_S = f_{\text{PIEZO}}(\upsilon, f). \tag{2}$$

The data of 42 experimental tests on a 5×5×36 mm^3 piezoelectric stack actuator are available to approximate and cross-validate f_{PIEZO}. υ has the values of 5,7.5,10,12.5,15 and 17.5 V, and the frequencies (f) are 20,30,40,50,60,70 and 80 Hz. In each experiment, with a pair of υ and f, R_S was tuned so that eventually met the aforementioned selection criterion. Ideally, such an R_S is the output of (2). The input-output data of 30, 6 and 6 randomly selected experiments were used as modelling, validation and test data sets, respectively. Fig. 3 shows that the available data of the second case study are quite dense. Having dense data does not contradict with small size of the data. In data-driven modelling, the inputs of a dense data set are distributed in the input space rather uniformly (Foster et al., 2021). Such a data set is small, if its size (number of samples) is small compared to the number of parameters of an appropriate model for such a problem.

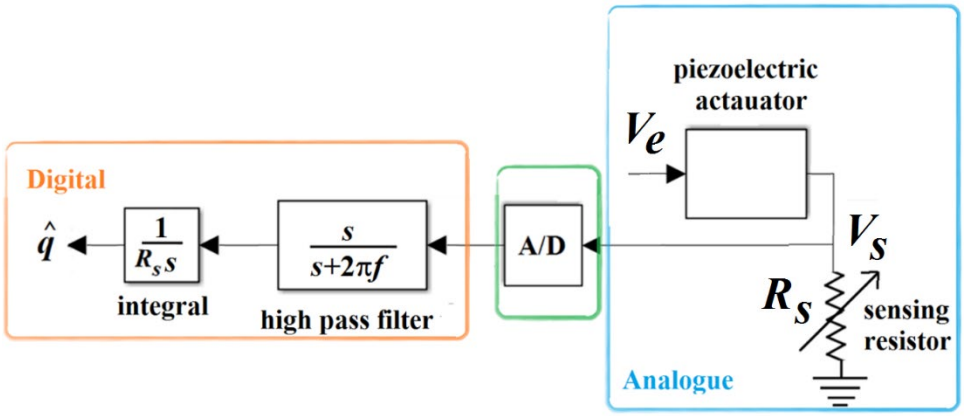

Figure 2. A schematic of a resistor-based charge estimator for a piezoelectric actuator

MODELLING

Development of a reliable data-driven model may include four tasks:

1- Mathematical Structure Definition
2- Parameter Identification
3- Overfitting Avoidance
4- Cross Validation

Up to three separate data sets, modelling, validation and test data, were used to perform the listed tasks for each problem defined in Problem Statement section. Generally speaking, the purpose of these quadruple steps is to minimise *error*, 'the discrepancy between the real output, from an experiment, and the estimated output by the model'. If the error is calculated using modelling, validation or test data, it is called modelling, validation or test error, respectively.

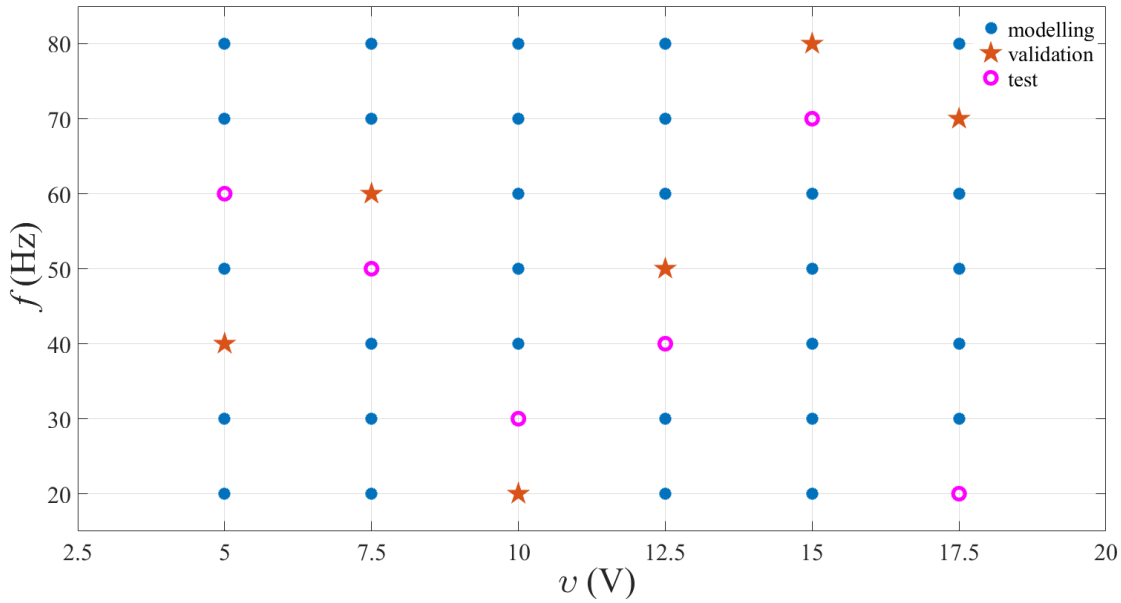

Figure 3. Distribution of modelling, validation and test data to approximate/identify and cross-validate f_{PIEZO}, only input space

Equation 3 mathematically defines the error in this research (Mohammadzaheri, Akbarifar, et al., 2020):

$$E = \frac{\sum_{t=1}^{n_d} (\hat{y}_t - y_t)^2}{n_d}. \tag{3}$$

where y is an output, n_d is the number of samples in the data set used to calculate the error and ^ refers to estimated values. Aforementioned quadruple tasks performed in data-driven modelling are briefly introduced in the following:

Mathematical Structure Definition: In some models, the mathematical structure is not certain from the beginning. For instance, in a neuro-fuzzy network (or in short, fuzzy model), the number of rules can be defined using the modelling data through subtractive clustering (Alibak et al., 2022), or in exact RBFNs, the size of the model depends on the modelling data (Mohammadzaheri, Emadi, et al., 2020).

Parameter Identification: Parameters of a data-driven model, with a known mathematical structure, are identified using the modelling data. Methods of parameter identification, generally, minimise the modelling error and have two categories: single-step and iterative methods. Some models, e.g. linear and RBFN models, use single-step identification methods such as non-recursive least square of error (LSE) (Mohammadzaheri et al., 2009; Saini et al., 2022). In iterative methods, e.g. the ones based on error propagation (Haykin, 1999), the parameters are tuned step by step to minimise the modelling error (also known as the training error, as detailed in Appendix A of (Mohammadzaheri, Tafreshi, et al., 2020)).

Overfitting Avoidance: Overfitting refers to excessive focus on decrease of the modelling error, which diminishes the generality of data-driven models (Cawley & Talbot, 2010; Mohammadzaheri et al., 2007). In iterative parameter identification, e.g. for MLPs, FCCs and neuro-fuzzy networks, at each iteration, the error is both calculated for the modelling and the validation data sets; while, the latter is not used for parameter identification. A discrepancy in trend of these dual errors (normally, increase of the validation error and ongoing decrease of the modelling error) is considered as a sign of overfitting and triggers to stop parameter

identification (Mohammadzaheri, Tafreshi, et al., 2020). In models with single-step parameter identifications, e.g. RBFNs, some specific parameters are identified with the validation data rather than with the modelling data to avoid overfitting (Morteza Mohammadzaheri et al., 2018a).

Cross Validation: Any data-driven model should fulfil the requirements of cross validation. In this paper, one round cross-validation or hold-out was employed, which requires that the estimation error of the model calculated with the test data (neither used in parameter identification nor in overfitting avoidance) is acceptable (Lendasse et al., 2003). In short, the test error should be reasonably low to cross validate a model. It should be noted that the validation data were not used to perform cross validation.

Six types of data-driven models were developed in this research to tackle problems detailed in Problem Statement section. In following subsections, a brief explanation of each model is presented with a focus on four aforementioned tasks for data driven modelling and correct use of the modelling and the validation data. All the developed models have a single output of y and n inputs of u_i, i=1,…,n.

Linear Models

In these models, the output is a linear combination of inputs

$$y = \sum_{j=1}^{n} \mathbf{A}_i u_i + \mathbf{A}_{i+1}. \tag{4}$$

Nothing needs to be done to define the mathematical structure of this model (i.e. task 1 in the list of quadruple tasks at the beginning of Modelling section), as the mathematical structure is evident. Model parameters (elements of **A**) were identified with single-step method of LSE (Mohammadzaheri et al., 2009). Overfitting was disregarded in development of (4) (i.e. task 3 was not performed); thus, both the modelling and the validation data were used for modelling.

Multi-layer Perceptrons (MLPs)

The employed MLPs have one hidden layer with m neurons and activation function of ϕ.

$$y = \sum_{j=1}^{m} \mathbf{B}_j \phi\left(\sum_{i=1}^{n} \mathbf{C}_{ij} u_i + \mathbf{D}_i \right) + \mathbf{D}_{i+1}, \tag{5}$$

where

$$\phi(x) = \frac{2}{1+\exp(-2x)} - 1. \tag{6}$$

MLPs, presented by (5) and (6), are universal approximators. That is, the model has a proven capability to model any system when sufficient data are available (Chen & Chen, 1995; Mohammadzaheri, Chen, et al., 2012) .

In this research, m=2n+1, (7), based on recommendation of (Haykin, 1999). Considering (7), the mathematical structure would be known. Nguyen-Widrow algorithm was used to suggest initial values for parameters (Nguyen & Widrow, 1990). Then, error back propagation with Levenberg-Marquardt algorithm (Mohammadzaheri & Chen, 2010) was utilised to minimise the modelling error iteratively and to identify MLP parameters. At each iteration, the validation error was calculated. Parameter identification stopped as the trend of the modelling and the validation errors became discrepant, i.e. overfitting happened. Even with use of parameter initialisation algorithms, some initial values of parameters may push the utilised parameter identification method to be trapped in local minima of the modelling error, leading

to low accuracy of the model (Mohammadzaheri et al., 2021). Consequently, parameter identification was repeated with different initial parameters. The model with the lowest validation error was chosen in the end.

Fully Connected Cascade (FCC) Networks

The employed FCC networks are very similar to the MLPs, with extra parameters (**E** elements) which connect the inputs directly to the output.

$$y = \sum_{j=1}^{n} \bar{\mathbf{B}}_j \phi\left(\sum_{i=1}^{m} \bar{\mathbf{C}}_{ij} u_i + \bar{\mathbf{D}}_i \right) + \sum_{i=1}^{n} \mathbf{E}_i u_i + \bar{\mathbf{D}}_{i+1}. \tag{8}$$

FCC networks have shown their merit in solving some non-engineering benchmarks (Hunter et al., 2012). The number of hidden layer neurons, m, was considered same as the one of MLPs, as the same recommendation of (7) is valid for FCC networks (Hunter et al., 2012). Parameter identification, overfitting avoidance and evasion from local minima of the modelling error in FCC networks are similar to the ones of MLPs.

Neurofuzzy Networks

Linear Sugeno-type fuzzy models were used in this research which are convertible to neuro-fuzzy networks (Ahmadpour et al., 2009). Such fuzzy models have k rules, each with n membership functions (one per input). For j^{th} rule and i^{th} input, the Gaussian membership function of (9) was employed to produce a membership grade, μ_{ij}, based on the input, u_i (Mehrabi et al., 2017):

$$\mu_{ij} = \exp\left(-\frac{(u_i - \mathbf{F}_{ij})^2}{2\mathbf{G}_{ij}^{\ 2}} \right). \tag{9}$$

The product of membership grades of a rule was considered as the weight of the rule, a real number between zero and one. The output of the whole model is the weighted sum of rule outputs (Mehrabi et al., 2017):

$$y = \frac{\sum_{j=1}^{k} \left(\overbrace{\left(\sum_{i=1}^{n} \mathbf{H}_{ij} u_i + \mathbf{I}_j \right)}^{j^{th}\ rule\ output} \prod_{i=1}^{n} \mu_{ij} \right)}{\sum_{j=1}^{k} \underbrace{\prod_{i=1}^{n} \mu_{ij}}_{j^{th}\ rule\ weight}}. \tag{10}$$

Neuro-fuzzy models, presented by (9) and (10), are universal approximators (Ying, 1998). The mathematical structure of the fuzzy model, e.g. the number of rules (k), was defined through subtractive clustering with use of the modelling data, the utilised subtractive clustering algorithm is similar to the one detailed subsection 2-3 of (Mohammadzaheri, Grainger, et al., 2012b).

Parameters were identified using an iterative method. At each iteration, gradient descent error back propagation algorithm was used to adjust elements of **F** and **G**, and LSE was used to adjust elements of **H** and **I** (Jang et al., 2006; M Mohammadzaheri et al., 2018). The validation error, calculated at every iteration, was used to stop parameter identification procedure and to avoid overfitting, in the same way as used for MLPs.

Radial Basis Function Netrworks

RBFNs, which are universal approximators too (Park & Sandberg, 1993), are presented as a combination of (11) and (12). They receive an array of inputs rather than inputs of a single data sample; a data sample has n inputs. An RBFN can estimate the output of maximum w data samples, where w is the number of data samples used to develop the model. If the input of fewer number of data samples, i.e. z, are fed into the model, first z columns of $\mathbf{O}$ and $\mathbf{L}$ are used.

$$\mathbf{O}_{ik} = \exp\left(-\left(S\underbrace{\sum_{j=1}^{n}\left(\mathbf{J}_{ij}-\mathbf{U}_{jk}\right)^2}_{\substack{\text{distance between input}\\ \text{and weight arrays}}}\right)^2\right). \tag{11}$$

$$\hat{\mathbf{Y}}_{1\times w} = \mathbf{K}_{1\times w}\times\mathbf{O}_{w\times w} + \mathbf{L}_{1\times w}. \tag{12}$$

(12) indicates that greater elements of $\mathbf{O}$ are more influential on the network's output. In addition, (11) shows that (i) the range of $\mathbf{O}$ elements is [0 1] and (ii) if the i^{th} row of $\mathbf{J}$ is identical to the k^{th} column of $\mathbf{U}$, then $\mathbf{O}_{ik}$ will be maximum, 1.

In RBFN modelling, arrays of $\mathbf{J}_{w\times n}$, $\mathbf{K}$ and $\mathbf{L}$ and the scalar of S namely 'spread' should be identified. At model development stage, where modelling data were used, (13) was used instead of (12). ^ is unnecessary in (13), since no estimation happens during model development:

$$\mathbf{Y}_{1\times w} = \begin{bmatrix}\mathbf{K} & \mathbf{L}\end{bmatrix}_{1\times 2w}\begin{bmatrix}\mathbf{O}\\ \mathbf{I}\end{bmatrix}_{2w\times w}. \tag{13}$$

In exact RBFNs, $\mathbf{J}=\mathbf{U_M}^{\mathrm{T}}$ (14), where $\mathbf{U_M}^{\mathrm{T}}$ is the transpose of an array of all inputs of the modelling data. Hence, w equals the number of modelling data samples, and the mathematical structure is identified. For instance, for the second problem of Problem Statement section, $\mathbf{U_M}^{\mathrm{T}}$ has the size of 2×30. Considering (11) and (14), $\mathbf{O}$ elements on the output, calculated with the modelling data, will be 1. Here is a pseudo-algorithm of exact RBFN modelling (to find $\mathbf{J}$, $\mathbf{K}$, $\mathbf{L}$ and S using the input and output arrays of the modelling data, $\mathbf{U_M}$ and $\mathbf{Y_M}$)

1. Set $\mathbf{J}= \mathbf{U_M}^{\mathrm{T}}$
2. Set $\mathbf{O}_{w\times w}$=ones($w\times w$)
3. Form and solve (13) with $\mathbf{Y_M}$ and $\mathbf{O}$ from step 2 to find $\mathbf{K}_{1\times w}$.
4. Find S , with trial and error, so as to minimise the validation error of the developed RBFN (anti-overfitting step)

An alternative to exact RBFN modelling is efficient RBFN modelling, which may produce RBFNs with fewer parameters. In this research, despite exact RBFNs, that employ the transpose of inputs array of the modelling data as $\mathbf{J}$, in efficient RBFN modelling, some columns of $\mathbf{U_M}$ were selected and transposed to form $\mathbf{J}$ (Morteza Mohammadzaheri et al., 2018b). Hence, the number of $\mathbf{J}$ rows, w, is smaller or equal to the number of the columns of $\mathbf{U_M}$, named $w_{\max}$ in this paper.

Prior to select $\mathbf{U_M}$ columns to be used as $\mathbf{J}$ rows, S, and a target error, E_t should be defined. For each set of S and E_t, every single column of $\mathbf{U_M}$ was transposed and tried as a single-row $\mathbf{J}$. Then, the corresponding RBFN was created using $\mathbf{K}$ and $\mathbf{L}$ calculated with (13). The column of $\mathbf{U_M}$ leading to the smallest modelling error was selected, transposed and used as the first row

of **J.** Afterwards, the remaining columns of **U** were examined to find the one in which addition of its transpose to **J** led to the largest drop in the modelling error**.** Transposed of such a column was added to **J**. This continued till the modelling error reached E_t. Thus, the mathematical structure of efficient RBFNs is defined with use of the modelling data. In this research, the entire process of finding **J** was repeated for different pairs of S and E_t, and the validation error was calculated for each pair.

Here is a pseudo-algorithm of efficient RBFN modelling:

1. **J**=null, $\mathbf{U}_{rem}$= $\mathbf{U}_{\mathbf{M}}$, $\mathbf{U}_{opt}$=null, E=VEX=1000 (a large number), $^{\mathbf{T}}\mathbf{J}$=null (temporary weight matrix)
2. Choose a large S and a target modelling error, E_t
3. Set w=1
4. Set k=1
5. Add transpose of k^{th} column of $\mathbf{U}_{rem}$ to **J** to form $^{\mathbf{T}}\mathbf{J}$
6. Set $\mathbf{O}_{w\times w}$=ones($w\times w$)
7. Solve (13) to find **K** and **L** ($\mathbf{Y}_{\mathbf{M}}$ and **O** are available from the modelling data and step 6)
8. Find the modelling error, ME. The model (11 and 12) needs to be rum more than once as $w< w_{max}$.
9. If $ME<E$, then E=ME and $\mathbf{U}_{opt}$=$\mathbf{U}_k$
10. k=k+1
11. If $k \leq (w_{max} - w+1)$ then go to 5
12. Remove $\mathbf{U}_{opt}$ from $\mathbf{U}_{rem}$ and add it to **J**
13. w=w+1
14. if $E>E_t$ then go to 4
15. Find the validation error , VE
16. If $VE<VEX$ then VEX=VE
17. If VEX is unacceptable go to 2

Choice of S and E_t was performed using full space search with zooming (use of smaller step-size) at low error areas.

Line 4 of exact RBFN modelling pseudo-codes and lines 14-17 of efficient RBFN modelling pseudo-codes use the validation data to tackle overfitting. Use of the modelling data at these lines would diminish generality of the model, and use of the test data would violate the conditions of cross validation.

Section Summary

Table 1 summarises the tasks performed in development of each model and the data used for each task. MD and VD refer to the modelling and the validation data, respectively. Two last columns refer to avoidance of overfitting through different strategies: (1) stopping parameter identification in the case of discrepancy in trend of modelling and validation errors, used for MLP, FCC and neuro-fuzzy networks, and (2) identifying some parameters with the validation data to improve generality of the models, or dual identification, used for RBFNs.

Table 1. Development stages for different models and their associated data

Model	Structure Definition	Parameter Identification	Over-fitting Avoidance-Stop Process	Over-fitting Avoidance- Dual Identification
Linear		MD+VD		
MLP		MD	VD	
FCC		MD	VD	
Fuzzy	MD	MD	VD	
RBFN	MD	MD		VD

RESULTS AND ANALYSIS

The models, reported in the previous section, were developed to approximate f_{ESP} and f_{PIEZO} introduced in Problem Statement section. For (neuro-) fuzzy modelling, subtractive clustering was performed with the influence range of 0.5 and squash factor of 1.25 for both problems. Also, accept ratios of 0.1 and 0.5 and reject ratios of 0.05 and 0.15 were used for subtractive clustering for the purpose of f_{ESP} and f_{PIEZO} approximation, respectively. The aforementioned factors have been explained in (Mohammadzaheri, Grainger, et al., 2012b). Exact RBFNs were developed with the spreads (S) of 76 and 41 for the first and the second problem, respectively. Efficient RBFNs were developed with S and E_t of 58 and 30 for the first problem (to approximate f_{ESP}) and 83.5 and 1.2 for the second problem. The results showed that targeting a too small modelling error (e.g. 0) increases the validation error or rises the chance of overfitting.

Tables 2 and 3 present different statistics for the developed models, all calculated with the test data, for the purpose of comparison of different techniques. MAE and MSE stand for mean of absolute error and mean of squared error. The range of output for the problems associated with Tables 2 and 3 are [28,72] ft and [17.5,225] Ω, respectively. The results evidently show that both systems are highly nonlinear, as linear models practically fail to approximate both f_{ESP} and f_{PIEZO}.

Table 2. Different statistics of eestimation error for different models to approximate f_{ESP}

	MLP	FCC	Fuzzy	RBFN Efficient	RBFN Exact	Linear
MAE	2.692	2.342	2.824	8.006	8.611	174.50
Error Bias	1.975	1.360	0.967	0.929	0.473	174.50
Error Variance	14.124	8.732	16.832	96.49	240.44	8484.3
MSE	18.025	10.582	17.767	97.35	240.66	38934
Number of Parameters	36	39	80	26	371	4

Table 3. .Different statistics of eestimation error for different models to approximate f_{PIEZO}

	MLP	FCC	Fuzzy	RBFN Efficient	RBFN Exact	Linear
MAE	0.623	1.386	2.652	1.537	2.000	54.916
Error Bias	0.127	1.265	0.370	1.070	-1.081	44.177
Error Variance	0.555	2.020	11.44	1.792	11.146	2682.8
MSE	0.571	3.620	11.58	2.937	12.315	4634.4
Number of Parameters	21	23	56	37	121	3

For the first case study (approximation of f_{ESP}, presented in Table 2), with fairly sparse data, FCC outperforms other models, with a sizably lower estimation error variance. This result is in agreement with the conclusions of (Hunter et al., 2012). The performance of MLP and Fuzzy models are fairly close to the FCC. MAE, a sensible criterion of accuracy, of the MLP is only around 15% larger than the one of FCC. Fuzzy and RBFN models which convert the parameter identification problem to a linear algebra problem, partly or in full, show small estimation biases of lower than 1 ft, almost 5 to 11 inches smaller than the bias of FCC head estimation.

MLP, however, shows evident superiority for the second case study (approximation of f_{PIEZO}, presented in Table 3) with dense data. MLP has the smallest estimation bias, estimation variance and number of parameters compared to all other nonlinear models. Alternatively, due to high density of the data, one can guess interpolation techniques may estimate the sensing resistance accurately in this case study. However, a similar research has shown that RBFN models outperform cubic interpolation and averaging methods (Mohammadzaheri, Emadi, et al., 2020). As a result, superiority of MLP can be also extended to these interpolation techniques.

In summary, for development of models to approximate nonlinear systems/functions with small data and for engineering purposes, three following recommendations can be drawn from the results of this research:

1. FCC is recommended to be employed in the case of sparsity of data; although, MLP and fuzzy models are also worth to be tried.
2. MLP is suggested to be employed with dense data.
3. RBFN models are not recommended, due to relatively high number of parameters and low accuracy. RBFNs were not the best models for any type of estimation purposes; the reason may be their inherited weakness against overfitting.

CONCLUSION

This chapter investigated the capability of a variety of common artificial intelligence techniques in data-driven modelling of engineering systems in the case of access to small data (small number of data samples). Five different AI models were even-handedly assessed in data-driven modelling of two case studies with sparse and dense data. Both systems were shown to be highly nonlinear.

For modelling with sparse data, FCC outperformed other techniques, closely followed by MLP and fuzzy models. This outcome is consistent with the literature claiming that FCCs are at an advantage over other AI modelling tools. However, for the problem with dense data, MLP showed an obvious superiority. RBFN models could not excel in any of the investigated data-driven modelling problems; therefore, they are recommended to be disregarded in data-driven modelling of engineering systems with small data. As a suggestion for future research, the presented comparison can be repeated for any other data-driven modelling technique not covered in this chapter or the techniques that will be developed in the future.

DISCLOSURE OF POTENTIAL CONFLICT OF INTEREST

The authors have no conflicts of interest, associated with this paper, to declare.

REFERENCES

Ahmadpour, M., Yue, W. L., & Mohammadzaheri, M. (2009). Neuro-fuzzy Modelling of Workers Trip Production. 32nd Australasian Transport Research Forum, Auckland, New Zealand.

Alibak, A. H., Alizadeh, S. M., Davodi Monjezi, S., Alizadeh, A. a., Alobaid, F., & Aghel, B. (2022). Developing a Hybrid Neuro-Fuzzy Method to Predict Carbon Dioxide (CO2) Permeability in Mixed Matrix Membranes Containing SAPO-34 Zeolite. *Membranes*, *12*(11), 1147.

Bazghaleh, M., Grainger, S., & Mohammadzaheri, M. (2018). A review of charge methods for driving piezoelectric actuators. *Journal of Intelligent Material Systems and Structures*, *29*(10), 2096-2104.

Bazghaleh, M., Grainger, S., Mohammadzaheri, M., Cazzolato, B., & Lu, T. (2013). A digital charge amplifier for hysteresis elimination in piezoelectric actuators. *Smart Materials and Structures*, *22*(7), 075016.

Castro, R. (2018). Data-driven PV modules modelling: Comparison between equivalent electric circuit and artificial intelligence based models. *Sustainable Energy Technologies and Assessments*, *30*, 230-238.

Cawley, G. C., & Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, *11*(Jul), 2079-2107.

Chang, C.-J., Li, D.-C., Huang, Y.-H., & Chen, C.-C. (2015). A novel gray forecasting model based on the box plot for small manufacturing data sets. *Applied mathematics and computation*, *265*, 400-408.

Chen, T. P., & Chen, H. (1995). Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks. *IEEE Transactions on Neural Networks*, *6*(4), 904-910. <Go to ISI>://A1995RF58200008

Collins, J., Brown, J., Schammel, C., Hutson, K., Jeffery, W., & Edenfield, M. (2017). Meaningful Analysis of Small Data Sets: A Clinician's Guide. *Clinical and Translational Research*, *2*(1), 16-19.

Foster, D., Gagne, D. J., & Whitt, D. B. (2021). Probabilistic Machine Learning Estimation of Ocean Mixed Layer Depth From Dense Satellite and Sparse In Situ Observations. *Journal of Advances in Modeling Earth Systems*, *13*(12), e2021MS002474.

Garg, A., Vijayaraghavan, V., Wong, C., Tai, K., Sumithra, K., Mahapatra, S., . . . Yao, L. (2015). Application of artificial intelligence technique for modelling elastic properties of 2D nanoscale material. *Molecular Simulation*, *41*(14), 1143-1152.

Goel, M., Sharma, A., Chilwal, A. S., Kumari, S., Kumar, A., & Bagler, G. (2023). Machine learning models to predict sweetness of molecules. *Computers in Biology and Medicine*, *152*, 106441.

Haykin, S. (1999). *Neural Networks A Comprehensive Introduction*. Prentice Hall, New Jersey.

Hunter, D., Yu, H., Pukish III, M. S., Kolbusz, J., & Wilamowski, B. M. (2012). Selection of proper neural network sizes and architectures—A comparative study. *IEEE Transactions on Industrial Informatics*, *8*(2), 228-240.

Jang, J. R., Sun, C., & Mizutani, E. (2006). *Neuro-Fuzzy and Soft Computing*. Prentice-Hall of India.

Kitchin, R., & Lauriault, T. P. (2015). Small data in the era of big data. *GeoJournal*, *80*(4), 463-475.

Kokol, P., Kokol, M., & Zagoranski, S. (2022). Machine learning on small size samples: A synthetic knowledge synthesis. *Science Progress*, *105*(1), 00368504211029777.

Lea, J. F., & Bearden, J. (1982). Effect of gaseous fluids on submersible pump performance. *Journal of Petroleum Technology*, *34*(12), 922-930.

Lendasse, A., Wertz, V., & Verleysen, M. (2003). Model selection with cross-validations and bootstraps—application to time series prediction with RBFN models. *Artificial Neural Networks and Neural Information Processing—ICANN/ICONIP 2003*, 174-174.

Li, W., Demir, I., Cao, D., Jöst, D., Ringbeck, F., Junker, M., & Sauer, D. U. (2022). Data-driven systematic parameter identification of an electrochemical model for lithium-ion batteries with artificial intelligence. *Energy Storage Materials*, *44*, 557-570.

Li, X., Jia, R., Zhang, R., Yang, S., Chen, G., & Safety, S. (2022). A KPCA-BRANN based data-driven approach to model corrosion degradation of subsea oil pipelines. *Reliability Engineering*, *219*, 108231.

Liu, X., Yan, Z., Wu, J., Huang, J., Zheng, Y., Sullivan, N. P., . . . Pan, Z. J. J. o. E. C. (2023). Prediction of impedance responses of protonic ceramic cells using artificial neural network tuned with the distribution of relaxation times.

Madni, A. M., & Sievers, M. (2018). Model-based systems engineering: motivation, current status, and needed advances. In *Disciplinary Convergence in Systems Engineering Research* (pp. 311-325). Springer.

Mehrabi, D., Mohammadzaheri, M., Firoozfar, A., & Emadi, M. (2017). A fuzzy virtual temperature sensor for an irradiative enclosure. *Journal of Mechanical Science and Technology*, *31*(10), 4989-4994.

Mohammadzaheri, M., Akbarifar, A., Ghodsi, M., Bahadur, I., AlJahwari, F., & Al-Amri, B. (2020). Health Monitoring of Welded Pipelines with Mechanical Waves and Fuzzy Inference Systems. International Gas Union Research Conference, Muscat,Oman,

Mohammadzaheri, M., AlQallaf, A., Ghodsi, M., & Ziaiefar, H. (2018). Development of a Fuzzy Model to Estimate the Head of Gaseous Petroleum Fluids Driven by Electrical Submersible Pumps. *Fuzzy Information and Engineering*, *10*(1), 99-106.

Mohammadzaheri, M., Amouzadeh, A., Doustmohammadi, M., Emadi, M., Nasiri, N., Jamshidi, E., . . . Soltani, P. (2021). Fault diagnosis of an automobile cylinder block with neural process of modal information. *International Journal of Mechanical and Mechatronics Engineering*, *21*(2), 1-8.

Mohammadzaheri, M., & Chen, L. (2010). Intelligent predictive control of a model helicopter's yaw angle. *Asian Journal of Control*, *12*(6), 667-679.

Mohammadzaheri, M., Chen, L., Ghaffari, A., & Willison, J. (2009). A combination of linear and nonlinear activation functions in neural networks for modeling a de-superheater. *Simulation Modelling Practice and Theory*, *17*(2), 398-407. https://doi.org/10.1016/j.simpat.2008.09.015

Mohammadzaheri, M., Chen, L., & Grainger, S. (2012). A critical review of the most popular types of neuro control. *Asian Journal of Control*, *16*(1), 1-11.

Mohammadzaheri, M., Emadi, M., Ghodsi, M., Bahadur, I. M., Zarog, M., & Saleem, A. (2020). Development of a Charge Estimator for Piezoelectric Actuators: A Radial Basis Function Approach. *International Journal of Artificial Intelligence and Machine Learning (IJAIML)*, *10*(1), 31-44.

Mohammadzaheri, M., Emadi, M., Ghodsi, M., Jamshidi, E., Bahadur, I., Saleem, A., & Zarog, M. (2019). A variable-resistance digital charge estimator for piezoelectric actuators: An alternative to maximise accuracy and curb voltage drop. *Journal of Intelligent Material Systems and Structures*, *30*(11), 1699-1705.

Mohammadzaheri, M., & Ghodsi, M. (2018). A Critical Review on Empirical Head-Predicting Models of Two-phase Petroleum Fluids in Electrical Submersible Pumps. *Petroleum & Petrochemical Engineering Journal*, *2*(4), 1-4.

Mohammadzaheri, M., Ghodsi, M., & AlQallaf, A. (2018a). Estimate of the head of the head produced by electrical submeersible pumps on gaseous petroleum fluids, a radial basis function network approach. *International Journal of Artificial Intelligence and Applications*, *9*(1), 53-62.

Mohammadzaheri, M., Ghodsi, M., & AlQallaf, A. (2018b). Estimate of the head produced by electrical submersible pumps on gaseous petroleum fluids, a raidal basis function netrwork approach. *International Journal of Artificial Intelligence and Applications*, *9*(1), 53-62.

Mohammadzaheri, M., Grainger, S., & Bazghaleh, M. (2012a). A comparative study on the use of black box modelling for piezoelectric actuators. *The International Journal of Advanced Manufacturing Technology*, *63*(9-12), 1247-1255.

Mohammadzaheri, M., Grainger, S., & Bazghaleh, M. (2012b). Fuzzy modeling of a piezoelectric actuator. *International Journal of Precision Engineering and Manufacturing*, *13*(5), 663-670.

Mohammadzaheri, M., Mirsepahi, A., Asef-afshar, O., & Koohi, H. (2007). Neuro-fuzzy modeling of superheating system of a steam power plant. *Applied Math. Sci*, *1*, 2091-2099.

Mohammadzaheri, M., Tafreshi, R., Khan, Z., Franchek, M., & Grigoriadis, K. (2015, 25 March 2015). *Modelling of Petroleum Multiphase Fluids in ESPs, an Intelliegnt Approach* Offshore Mediternean Conference, Ravenna, Italy.

Mohammadzaheri, M., Tafreshi, R., Khan, Z., Franchek, M., & Grigoriadis, K. (2016). An intelligent approach to optimize multiphase subsea oil fields lifted by electrical submersible pumps. *Journal of Computational Science*, *15*, 50-59.

Mohammadzaheri, M., Tafreshi, R., Khan, Z., Ghodsi, M., Franchek, M., & Grigoriadis, K. (2020). Modelling of petroleum multiphase flow in electrical submersible pumps with shallow artificial neural networks. *Ships and Offshore Structures*, *15*(2), 174–183.

Mohammadzaheri, M., Ziaiefar, H., & Ghodsi, M. (2022). Digital Charge Estimation for Piezoelectric Actuators: An Artificial Intelligence Approach. In *Handbook of Research on New Investigations in Artificial Life, AI, and Machine Learning* (pp. 117-140). IGI Global.

Mohammadzaheri, M., Ziaiefar, H., Ghodsi, M., Bahadur, I., Zarog, M., Saleem, A., & Emadi, M. (2019). Adaptive Charge Estimation of Piezoelectric Actuators, a Radial Basis Function Approach. 20th International Conference on Research and Education in Mechatronics Wels, Austria.

Mohammadzaheri, M., Ziaiefar, H., Ghodsi, M., Emadi, M., Zarog, M., Soltani, P., & Bahadur, I. (2022). Adaptive Charge Estimation of Piezoelectric Actuators with a Variable Sensing Resistor, an Artificial Intelligence Approach. *Engineering Letters*, *30*(1), 193-200.

Nguyen, D., & Widrow, B. (1990, 1990). *Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights* International Joint Conference on Neural Networks, San Diego, USA.

Pal, A., Zhu, L., Wang, Y., & Zhu, G. (2022). Data-driven model-based calibration for optimizing electrically boosted diesel engine performance. *International Journal of Engine Research*, *in press*, 14680874221090307.

Park, J., & Sandberg, I. W. (1993). Approximation and radial-basis-function networks. *Neural computation*, *5*(2), 305-316.

Rahbar, A., Mirarabi, A., Nakhaei, M., Talkhabi, M., & Jamali, M. (2022). A comparative analysis of data-driven models (SVR, ANFIS, and ANNs) for daily karst spring discharge prediction. *Water Resources Management*, *36*(2), 589-609.

Saini, S., Orlando, M. F., & Pathak, P. M. (2022). Intelligent Control of Master-Slave based Robotic Surgical System. *Journal of Intelligent Robotic Systems*, *105*(4), 1-20.

Steyerberg, E. W., Eijkemans, M. J., Harrell Jr, F. E., & Habbema, J. D. F. (2000). Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Statistics in medicine*, *19*(8), 1059-1079.

Taajobian, M., Mohammadzaheri, M., Doustmohammadi, M., Amouzadeh, A., & Emadi, M. (2018). Fault diagnosis of an automobile cylinder head using low frequency vibrational data. *Journal of Mechanical Science and Technology*, *32*(7), 3037-3045.

Ying, H. (1998). *General Takagi-Sugeno fuzzy systems are universal approximators*. <Go to ISI>://000074668800145

Zhang, T., Chen, J., Li, F., Zhang, K., Lv, H., He, S., & Xu, E. (2022). Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions. *ISA transactions*, *119*, 152-171.