Journal of CO2 Utilization 70 (2023) 102452

Contents lists available at ScienceDirect

Journal of CO2 Utilization

journal homepage: www.elsevier.com/locate/jcou

Predicting overall mass transfer coefficients of CO_2 capture into monoethanolamine in spray columns with hybrid machine learning

Ulderico Di Caprio [a], Min Wu [a], Florence Vermeire [b], Tom Van Gerven [c], Peter Hellinckx [d], Steffen Waldherr [b], Emine Kayahan [a,*,1], M. Enis Leblebici [a,*,1]

[a] Center for Industrial Process Technology, Department of Chemical Engineering, KU Leuven, Agoralaan Building B, 3590 Diepenbeek, Belgium
[b] Chemical Reactor Engineering and Safety (CREaS), Department of Chemical Engineering, KU Leuven, Celestijnenlaan 200F, B-3001 Heverlee, Belgium
[c] Process Engineering for Sustainable Systems, Department of Chemical Engineering, KU Leuven, Celestijnenlaan 200F, B-3001 Heverlee, Belgium
[d] Faculty of Applied Engineering, University of Antwerp, Groenenborgerlaan 171, 2000 Antwerp, Belgium

ARTICLE INFO

Keywords:
Hybrid modelling
CO_2 capture
Machine learning
Process intensification

ABSTRACT

In order to avoid the catastrophic effects of global warming, we need to reduce CO_2 emissions. Currently, the most mature technology to reduce large industrial CO_2 emissions is the absorption of CO_2 into aqueous monoethanolamine (MEA) solutions. The process is mostly studied in packed columns, for which many correlations have been offered to predict overall mass transfer coefficients (K_Ga). Spray columns are less prone to corrosion and were shown to enhance K_Ga. However, to the best of our knowledge, there are no models to predict K_Ga in spray columns. Hybrid modelling tools, a combination of machine learning techniques and first-principle information, showed remarkable capabilities in modelling complex systems. In this work, we applied hybrid modelling techniques benchmarking performances of four regressors: Ridge regression, decision tree regressor (DTr), support vector machine regressor (SVMr) and fully connected artificial neural network (ANN). We compared the performances of these modelling techniques with a model developed using the Buckingham Π-theorem, which is the most used state of the art technique to model K_Ga based on dimensionless numbers. SVMr and DTr showed higher accuracies among the trained models on the test set. SVMr can predict K_Ga within 6.4% error on the test set, whereas the Buckingham modelling approach resulted in 83 % error. The use of machine learning techniques resulted in predictive models with higher accuracies compared to the Buckingham Π-theorem. Predicting K_Ga with a higher accuracy allows more control over operational parameters and better column designs.

1. Introduction

Global warming is one of the main challenges that humanity has to tackle in the near future. In its sixth assessment report, the Intergovernmental Panel on Climate Change highlighted the urgency of actions to reduce the emissions of greenhouse gases (GHG) such as CO_2 [1]. Several solutions are available to reduce CO_2 footprint of chemical and energy industries, such as electrochemical separation, oxyfuel technologies, pre-combustion capture and post-combustion capture [2]. The latter requires little modification to currently available processes [3]. In this frame, the most mature technology for post-combustion CO_2 capture is the absorption of CO_2 into aqueous monoethanolamine (MEA) solutions [3,4] in packed or spray columns. The absorption process is mostly studied in packed columns on an industrial scale [4–6].

CO_2 absorption into amine solutions is complicated due to the interaction among kinetics, heat and mass transfer, hydrodynamics and thermodynamics [7,8]. The mass-transfer coefficient (K_Ga) is one of the most critical design parameters for absorption columns. The higher the K_Ga value, the lower the capital and operational costs of the process. Wang et al. reviewed correlations of the K_Ga for packed columns. Most of the K_Ga correlations in literature are empirical or semiempirical, obtained by experimental data fits [8]. Most of them rely on correlations between dimensionless numbers (i.e. ratios and products of process variables) to predict K_Ga [9]. The lack of understanding of the complex transport phenomena in packed reactors makes it challenging to evaluate the performance of the column under different experimental

* Corresponding authors.
E-mail addresses: emine.kayahan@kuleuven.be (E. Kayahan), muminenis.leblebici@kuleuven.be (M.E. Leblebici).
[1] Equal contribution

https://doi.org/10.1016/j.jcou.2023.102452
Received 16 December 2022; Received in revised form 19 February 2023; Accepted 6 March 2023
Available online 11 March 2023
2212-9820/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

conditions [7]. Razi et al. performed a rate-based simulation to assess the effect of several mass transfer coefficients and interfacial area on the system performance for the case of CO_2 capture with an aqueous MEA solution. They concluded that mass transfer correlations for the packed bed reactors have a significant degree of uncertainty in determining the column height due to the complex relationship between transport phenomena, mass transfer coefficients, interfacial areas and kinetics [10]. Recently, investigation of absorbent mixtures without water (water-lean solvents) has been gaining popularity because of their higher transfer coefficients and lower regeneration cost [11]. A significant contribution in this frame was performed by Wanderley et al. [12]. They investigated in-silico physical characteristics of water-lean solvent formulation to increase overall process efficiency. Their investigations suggest valuable guidelines for solvent selection. However, the study focuses only on partial characteristics of the system (i.e., process variables or solvent properties) without analysing their synergistic effects. For this reason, the authors indicated that this process could be further analysed by using statistical methods so that other optimisation parameters can be obtained.

Recently, some studies investigated the CO_2 capture process in spray columns. They showed that the K_Ga in spray columns could improve significantly compared to packed columns [13–15]. Solvents used to absorb CO_2 are mostly very corrosive [16]. Due to the lack of column internals, spray columns are less prone to corrosion and have lower pressure drops. Despite these advantages, to our best knowledge, a predictive model for K_Ga in spray columns for CO_2 capture is not available. To the best of our knowledge, the only study in the literature was a previous work from our research group. It employed a structure-optimised artificial neural network to predict the K_Ga, but it focused on the methodology rather than the case. In addition, it focused on a narrow MEA concentration range and only pressure nozzles [17]. There are some models to predict K_Ga for other applications in spray columns by using dimensionless numbers employing Buckingham Π-theorem. Rajmohan et al. predicted removal efficiency in a spray column for SO_2 capture by correlating dimensionless numbers describing the system [18]. The model achieved a coefficient of determination (R^2) of 0.99 and a root-mean-squared error (RMSE) of 0.0394. Another application of this technique was to predict K_Ga for a dehumidification column [19]. The final model has an average error of 10.4 %. Although these models can successfully predict K_Ga within an error range, they only apply within a specific range of operational conditions. As mentioned above, more advanced statistical tools are needed to predict K_Ga more accurately for a wider range of operational parameters due to the complexity of the overall systems.

Over the last decades, artificial intelligence (AI) and machine learning (ML) techniques have been applied to several fields of science and technology to boost the investigation and prediction capabilities of physical phenomena. These techniques develop highly accurate models using a large amount of historical data. ML approaches have already been applied to complex and nonlinear engineering problems [20,21]. ML techniques have been widely used for predictive tasks in the field of CO_2 capture [22–25]. Focusing on the post-combustion capture through absorption, ML techniques have been applied to predict CO_2 solubility in various liquid formulations [26–28]. Some efforts were also performed to predict K_Ga in packed columns employing MEA for CO_2 capture. Fu et al. used back-propagation (BPNN) and radial basis function (RBFNN) neural networks to model mass flux per unit of time from gas to liquid using process variables [29]. The model outperformed the state-of-the-art Onda correlation that partially utilises dimensionless numbers to predict the K_Ga value [30]. Unfortunately, the paper [29] does not report the final structures of the adopted neural networks; thus, evaluating their complexity and the correlation between the variables is impossible. However, increased performances of the RBFNN indicate a strong nonlinearity between the inputs and outputs of the system. Despite several studies focused on applying ML techniques to CO_2 capture, they have never been applied to predict K_Ga for CO_2 capture in a spray column. However, ML techniques were applied in spray columns for SO_2 capture. Valera et al. [31] modelled SO_2 capture by employing a BPNN. The model outputs were SO_2 removal percentage and volumetric mass transfer coefficients. The model returned an average error of 4.5 % on K_Ga. It utilised a simple network containing only one hidden layer with nine neurons. The inputs of the network were five process variables. However, this model was trained with only one dataset from the same column. Therefore, the generalisation capabilities of the model on other column configurations cannot be assessed. This makes the model highly case-specific, and its utilisation is limited to one specific column.

Usually, complex ML model structures are required to approximate a complex behaviour. On the other hand, a complex model structure requires a high amount of data with high variability to generalise the model prediction to new cases. However, some processes have limited available data, such as CO_2 capture in spray columns. This hinders prediction accuracies of the classical ML modelling approaches. In such cases, hybrid models can increase the model accuracy while keeping a reduced amount of data points. They integrate first-principle information (e.g., mass and heat balances) into statistical learning. This approach decreases the amount of data required for the training since more information is associated with each data point [32]. In addition, the hybrid methodology increases extrapolation capabilities of the model since it also relies on first-principle equations.

To our best knowledge, no papers are available in the literature on modelling K_Ga in a spray column for CO_2 capture. We intend to fill this gap. The aims of this paper are as follows: (1) to investigate the impact of process parameters on K_Ga, (2) to investigate the most appropriate ML technique to develop a reliable hybrid model of the process and (3) to develop an operative model for optimisation of spray columns. We performed a statistical analysis on process variables to evaluate the most impactful parameter on K_Ga. Then, these process parameters were correlated to K_Ga using hybrid modelling techniques. We trained and analysed four modelling techniques: Ridge regression, decision tree regressor (DTr), support vector machine regressor (SVMr) and a fully connected artificial neural network (ANN). The prediction accuracy and generalisation capabilities of each regressor were analysed through the coefficient of determination (R^2), mean squared error (MSE) and median absolute percentage error (MedAPE). The results were compared to the performance of a model obtained with dimensionless numbers characterising the system utilising the Buckingham Π-theorem. The final model obtained in this paper correlated the process parameters with K_Ga. It is suitable for in-silico optimisation because of its high accuracy.

In this work, the word “feature” is utilised to express the inputs of the model, such as gas flowrate and gas concentration, and the word “parameter” is utilised to indicate the internal parameters of the model that must be identified by the optimiser during the training phase.

2. Methods

2.1. Dataset description

The dataset to train the hybrid models was curated from six studies available in the literature (Table 1). The information reported in the studies were: (1) CO_2 concentration in gas phase at the inlet of the absorption column, (2) volumetric gas flow rate at the inlet of the absorption column, (3) MEA mass fraction in the absorbent liquid, (4) initial CO_2 concentration in the absorbent liquid, (5) volumetric liquid flow rate, (6) nozzle type, (7) absorption column diameter, (8) absorption column height and (9) temperature of the absorbent liquid at the column inlet. Some of the studies explicitly reported K_Ga values, while others reported a molar fraction of CO_2 at the outlet of the absorption column. For those that missed K_Ga information, it was calculated using the equation by Kayahan et al. [15] reported in Eq. (1).

$$K_Ga = \frac{G_{inert} \bullet (Y_{CO_2,in} - Y_{CO_2,out})}{Z \bullet P \bullet \left(y_{CO_2} - y^*_{CO_2}\right)_{lm}} \quad (1)$$

Table 1
Main information about the experimental data used to train the hybrid models in this work.

Reference	Nozzle	Data points	MEA mass fraction [%wt]	Surface area [$m^2 \cdot m^{-3}$]	$K_G a$ [$kmol \cdot m^3 \cdot h^1 \cdot kPa$]	
					Min	Max
Kuntz and Aroonwilas[13]	Pressure	173	18.3 – 30.4	113–1700[a]	6.20e-2	6.1
Wu et al.[14]	Pressure	16	10.0 – 35.0	168–200	0.30	1.70
Tamhankar et al.[33]	Pressure	18	30.0 – 40.0	n/a†	9.25e-3	0.148
Zimmermann et al.[34]	Pressure	25	30.0	n/a**	1.90e-3	5.52e-2
Seyboth et al.[35]	Pressure	25	30.0	n/a**	1.69e-3	4.71e-2
Kayahan et al.[15]	Ultrasonic	94	30.0 – 100	45.3–93.8	1.58	11.72
Final dataset	General	351	10 – 100	n/a	1.69e-3	11.72

† The authors measure and report the surface area of the spray; however, it cannot be used for the mass-transfer calculations and compared with the other values included in the table.
**The authors report the D_{32} expected by the nozzle; however, they do not give the number concentration of the droplets. Therefore, the surface area is impossible to estimate.
[a] Calculated using the K_G definition and empirical correlations.

where G_{inert} is the inert gas molar flux [$kmol \cdot m^{-2} \cdot h^{-1}$], P is the total pressure within the column [kPa], Z is the column height [m], Y_{CO_2} is the molar ratio of CO_2 in the gas phase, y_{CO_2} is the molar fraction of the CO_2 in the gas phase [mol/mol] and $y^*_{CO_2}$ is the molar fraction of CO_2 in the gas phase at the equilibrium. This value is assumed to be zero due to the fast reaction between MEA and CO_2 [7]. The subscript *lm* denotes the logarithmic mean between the inlet conditions and the outlet conditions of the column.

The study from Kuntz and Aroonwilas [13] reported the $K_G a$ value as it is utilised in this work. Wu et al. [14] reported the value of $\Phi = K_G a * (P_A - P^*_A)$ where P_A is partial pressure of CO_2 and P^*_A is the equilibrium concentration of CO_2 with the liquid. In this study, to compute $K_G a$ from [14], we considered partial pressure of CO_2 at the outlet of the column. In addition, the equilibrium concentration of CO_2 is set to be 0 because of the fast reaction between MEA and CO_2 [7]. All the other datasets were calculated using Eq. (1).

2.2. Hybrid modelling structure

The available data were expanded using physical information about the nozzle type and the absorbing liquid to increase generalisation capabilities of the regressor. Namely, (1) density of the absorbing liquid, (2) viscosity of the absorbent liquid, (3) surface tension of the absorbent liquid, and (4) Sauter mean diameter of droplets (D_{32}) were used aside from the experimental information reported in the source papers (Fig. 1). The model output was $K_G a$. Further information about the distribution of the dataset employed in this work and the computed physical information is reported in the Appendix of this work in sections A1, A2 and A3.

From Fig. 1, it is possible to notice that the surface area of the spray has not included in the input space of the models. This is because some papers used as data sources do not report a volumetric surface area [33] or do not report any information about the surface area [34,35]. In addition, the papers did not report any information to calculate the volumetric surface area (e.g., the particles number concentration within the spray). In addition, the volumetric surface area could have a high uncertainty when reported because of the methodology used for its estimation [13]. For all these reasons, the model did not include the volumetric surface area in the input space, but it contains variables from which it is possible to infer this information (i.e., liquid flow rate, gas flow rate, Sauter mean diameter, reactor geometry and physical properties of the mixture).

2.3. Pre-processing

The experimental and the first-principle physical variables were pre-processed before they were used in the regressors. Two operations were executed on them: computation of monomial and rational interactions and normalisation (Fig. 1).

2.3.1. Monomial and rational features computation and selection

Monomial and rational interactions between the input variables were computed to augment the information available to the model. A fraction of these was selected as model input. The monomial interactions had a maximum overall order of three and a maximum individual order of single variables of two. For example, considering the input variables A and B, the applied procedure returns the following monomials: A, B, $1/A$, $1/B$, AB, A/B, B/A, A^2B, B^2A, A^2/B, A/B^2, B^2/A, B/A^2.

After calculating the monomial and rational features, p-values among each input variable and $K_G a$ were computed. All the variables

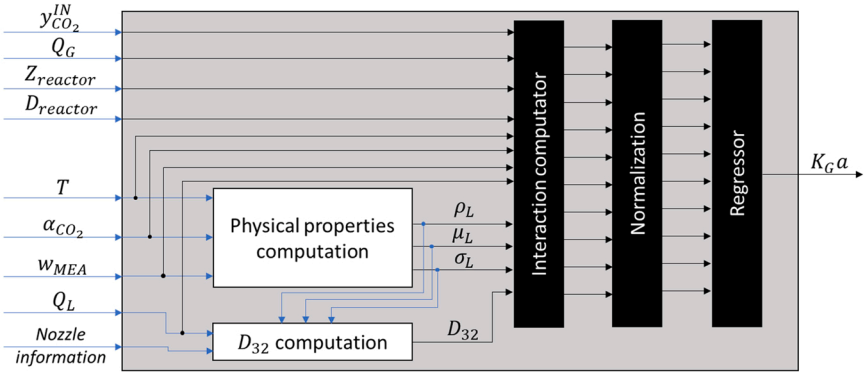

Fig. 1. Structure of the hybrid model and data flow in it. Part of the input information is utilised to compute physical properties of absorbent liquid and Sauter mean diameter (D_{32}) of droplets. This information and the experimental information were utilised to compute the most important interaction to use as input of the regressor. It returns predicted value of mass-transfer coefficient ($K_G a$). In this figure $y^{IN}_{CO_2}$ is molar fraction of CO_2 within flue gas [mol/mol], Q_G is flue gas flow rate [$m^3 \cdot h^{-1}$], $Z_{reactor}$ is height of the column [m], $D_{reactor}$ is diameter of the column [m], T is temperature [K], α_{CO_2} is concentration of CO_2 within the absorbent liquid at column inlet [mol $CO_2 \cdot mol^{-1}$ MEA], w_{MEA} is mass fraction of MEA within the absorbent liquid at the inlet of the column [kg/kg], Q_L is the flow rate of the absorbent liquid [$m^3 \cdot h^{-1}$], ρ_L is density of the absorbent liquid [$kg \cdot m^{-3}$], μ_L is viscosity of the absorbent liquid [Pa·s], σ_L is surface tension of the absorbent liquid [$N \cdot m^{-1}$], D_{32} is Sauter mean diameter of droplets [m], $K_G a$ is mass-transfer coefficient [$kmol \cdot m^{-3} \cdot kPa^{-1} \cdot h^{-1}$].

with a $p < 0.05$ were selected. This operation returned a list of variables with a high statistical influence on K_Ga; however, some were still linearly dependent. To remove this dependence, the Pearson correlation coefficient (ρ_{X_1,X_2}) [36] was calculated for every couple of features; one of the two members of the couple having with $\rho_{X_1,X_2} > 0.85$ were eliminated.

At the end of this operation, we identified 27 independent features with a high statistical dependence on K_Ga. These features were used as inputs of the statistical regressor. The reader can find a list of the features used as inputs of the regressor in the Appendix of this work in section A4.

2.3.2. Normalisation

After the computation and selection of the monomial and rational features, we normalised the experimental values of the input features of the statistical regressor and K_Ga values. This procedure is required to increase the quality of the model training. The scaling is executed as reported in Eq. (2).

$$\boldsymbol{x}_i^{sc} = \frac{\boldsymbol{x}_i - \min(\boldsymbol{x}_i)}{\max(\boldsymbol{x}_i) - \min(\boldsymbol{x}_i)} \tag{2}$$

where $\boldsymbol{x}_i^{sc}$ is the normalised value of the variable $\boldsymbol{x}_i$. The value of each feature is rescaled using minimum and maximum values that the variable can have across the dataset. This procedure returns values of the scaled variables in the range [0,1].

2.4. Cross-validation

Cross-validation is a common practice in statistical modelling. It is needed to estimate the interpolation accuracy and generalisation capabilities of the model to avoid overfitting training data. The "Single hold-out random subsampling" is a more straightforward technique, where the model results are evaluated on data contained in a single test dataset. The test points are often randomly selected, meaning they are in the same space as the training set but cover different scenarios. Often the test set is a part of the original dataset. Therefore, using this methodology, the original dataset is split into two parts: one is used for training the model, while another is used to test the model generalisation capabilities [37]. This approach is simple to implement but excludes a part of the dataset from the training. This can be problematic when the dataset available for the training is small, like the case in this study. A promising technique which can use the entire dataset for model training while still allowing estimation of the generalisation capabilities of the model is k-fold cross-validation. With this technique, the dataset is divided into *k* parts, one of these parts is chosen as the test set, and the other *k-1* parts are chosen as the training set. A model is trained and subsequently tested on this split. The procedure is repeated until all parts served as a test set. The estimation of the model accuracy on new data and generalisation capabilities are an average of the accuracy of each model [38]. The value of k is chosen in the design phase. In literature, the best solution was found by taking k equal to the amount of data within the dataset. In other words, every data point has to serve as the test set alone. However, this solution is computationally expensive because it requires as much model training as the number of data points in the dataset. Therefore, lower values of k are often used. Several studies suggest that $k = 10$ can be a good approximation [38–40].

In this study, a hybrid approach to perform the model cross-validation was applied. The original dataset containing 351 data points was split into two parts. The training set contained 311 data points, which were used to perform the training and the identification of hyperparameters of the regressors employed in this study. The identification of the ML models' hyperparameters was performed by using 10-fold cross-validation executed only on the training set. The hyperparameters that returned a higher score on 10-fold cross-validation were chosen. Subsequently, training of the model was performed by using the entire training set. The generalisation capabilities of the model were verified through the test set. It was composed of the remaining 40 data points of the original dataset, which were not contained in the training set.

2.5. Modelling approaches

This work tested and analysed the prediction and generalisation performances of four regression techniques. Namely, they are Ridge regression, DTr, SVMr and ANN. They have already shown high prediction accuracies in chemistry and chemical engineering tasks [41–44]. All models were trained using the scikit-learn (version 0.24.2) library in Python [45]. In the following paragraphs, each regressor is discussed in detail.

2.5.1. Ridge regression

Ridge regression is a formulation of linear regression. Linear regression is a statistical approach to compute the output of a system based on its input values. This task is performed through the identification of model parameters. Each parameter is related to one input feature. The model training aims to search the set of model parameters that minimises errors between experimental output values and model predictions. An optimisation algorithm performs the search for the optimal parameters set. The final structure of the linear model is reported in Eq. (3).

$$\bar{y}(\boldsymbol{x}) = \sum_{i=1}^{N_{features}} \omega_i x_i + \gamma \tag{3}$$

where $\bar{y}(\boldsymbol{x})$ is the value of the model output, which is related to the input value $\boldsymbol{x}$, ω_i is the *i*th component of the vector $\boldsymbol{\omega}$ that contains all the parameters, x_i is the *i*th component of the input array $\boldsymbol{x}$, γ is the value of the constant bias.

The optimiser aims to identify the values of $\boldsymbol{\omega}$ and γ to decrease the error $|\bar{y}(\boldsymbol{x}) - y(\boldsymbol{x})|$. In this paper, the L2 regularisation technique has been applied to avoid overfitting training data because of the small number of experiments in the dataset. A Ridge regression was applied to execute the modelling task [46]. Thus, the function that the optimiser aims to minimise is:

$$\sum_{i=0}^{N_{examples}} \left| y_i - \left(\sum_{j=1}^{N_{features}} \omega_j x_{i,j} + \gamma \right) \right|^2 + \lambda \sum_{i=0}^{N_{parameters}} \omega_i^2 \tag{4}$$

where y_i is the output associated with the input array $\boldsymbol{x}_i$, ω_i is the *i*th component of the vector $\boldsymbol{\omega}$ that contains all the parameters, $x_{i,j}$ is *j*th experimental input associated with the *i*th experimental output, and λ is the L2 regularisation coefficient. In this study, a grid search of the hyperparameter λ was performed following a 10-fold cross-validation strategy using only the training dataset. The search found that $\lambda = 0.0559$ retuned higher R^2 on 10-fold cross-validation.

2.5.2. Decision tree regressor (DTr)

Unlike Ridge regression, DTr does not utilise continuous functions containing tuneable parameters to perform a prediction task. On the contrary, DTr is composed of a decisional structure with several branches. Prediction task consists of selecting a proper branch according to input conditions. Structure of the tree can have several levels to refine the decision. When the decision has been taken at the first level of the tree, the information is moved to the second level, where another decision is taken, and so on, until arriving at the deepest level representing the final answer to the problem. This structure is highly suitable for classification tasks. Several algorithms are available to train a decision tree, and all utilise the minimisation of metrics such as entropy of the information within various decision groups [47]. The training is based on building several decision structures. Then, the one with the lowest value of the metric is chosen as the final model [48]. The decision tree

structure can also be applied to regression tasks. In this case, a split between the nodes is done to decrease the error between experimental and modelled output [49]. In this work, a grid search of the depth of the tree was performed by a 10-fold cross-validation strategy using only the training dataset. The search found that a depth of 8 levels returned the highest R^2 on the 10-fold cross-validation. Mean squared error was the criterion utilised to evaluate quality of the split during the training procedure.

2.5.3. Support vector machine regressor (SVMr)

Like the DTr, the SVMr was initially developed to accomplish classification tasks rather than regression. They have been broadly used in the literature for this purpose [50]. The peculiarity of this technique is to transform the input dataset containing nonlinear correlations into a dataset that can be divided using a hyperplane (i.e., a linear function). This operation is done by using transformation functions called kernels. They map the input dataset into a higher dimensional space that permits linear separation of variables [51]. A modification of SVMr allows regression and prediction of continuous output [52]. The training of an SVM regressor differs from the other techniques reported in this work. Usually, the model training aims to minimise the error between predictions and experimental values. On the contrary, in the SVMr, the training is performed to have all predictions within two boundaries in which predictions are considered acceptable. The width of this bend is ε. The training aims to tune the model parameters to reduce the number of points outside this band. This training approach allows SVMr to have excellent generalisation capabilities and perform well with small datasets [53]. The mathematical form for output value prediction is reported in Eq. (5).

$$\bar{y}(\boldsymbol{x}) = \sum_{i=1}^{N_{features}} \alpha_i * K(\boldsymbol{x}, \boldsymbol{x}_i) + b \tag{5}$$

$$K(\boldsymbol{x}, \boldsymbol{x}_i) = e^{-\gamma \|x_i - x\|^2} \tag{6}$$

where $\bar{y}(\boldsymbol{x})$ is the predicted output value, α_i is the coefficient associated with the ith feature, $\boldsymbol{x}_i$ are the support vectors, $K(\boldsymbol{x}, \boldsymbol{x}_i)$ is the vector returned by the chosen kernel associated with vectors $(\boldsymbol{x}, \boldsymbol{x}_i)$, and b is the intercept of the regression. In this work, kernel type and hyperparameters C, ε and γ of the SVR algorithm were chosen using a cross-validated grid search. The algorithm returned the model with the highest cross-validation R^2 using a radial basis function (RBF) kernel. The RBF equation is reported in Eq. (6). The following parameters were chosen from the grid search: C= 25.12, ε = 3.981e-3, γ = 0.1585.

2.5.4. Fully connected artificial neural network (ANN)

ANNs are both used for regression and classification tasks. They are based on an elementary structure called perceptron (Fig. 2a). Here, inputs are first linearly combined, and then the result of this operation is used as input for an activation function. The activation function in this structure aims to add nonlinearity to the system. Only tuneable parameters within neurons are coefficients of linear combination. The activation function is fixed during the design phase (i.e., it is a hyperparameter of the neuron). The transformation performed by each neuron can be formalised as reported in Eq. (7).

$$\bar{y}_{neuron} = \delta\left(\sum_{i=1}^{N} w_i * x_i + b\right) \tag{7}$$

$\bar{y}_{neuron}$ is the output of the neuron, δ is the activation function used by the neuron, N is the number of inputs fed to the neuron, w_i is the weight related to the ith feature of the input x_i and b is the bias associated with the neuron. Within the network, the neurons are organised in layers. Each layer contains a certain number of neurons. Neurons in a specific layer of the network use the information coming from the neurons of the layer before them as input and return their output to the neurons of the layer after them (Fig. 2b).

The number of layers and number of neurons per layer are two crucial network hyperparameters that must be optimised and determined prior to training the network [54]. Because of its modularity, this modelling approach has been widely used in the literature for regression and classification tasks [55]. For the same reason, ANNs have been widely employed for chemical engineering applications [56].

The training task for an ANN involves identifying weights and biases in the linear structure of each network neuron. This is performed using the back-propagation algorithm in which parameters of the network are tuned, taking into account the error between the predicted output and experimental output and its first derivative with the model parameters [57]. For this study, an ANN with one hidden layer containing 116 neurons has been trained using the Adam optimiser [58]. The number of optimal neurons within the hidden layer was found using a grid search with 10-fold cross-validation executed on the training dataset.

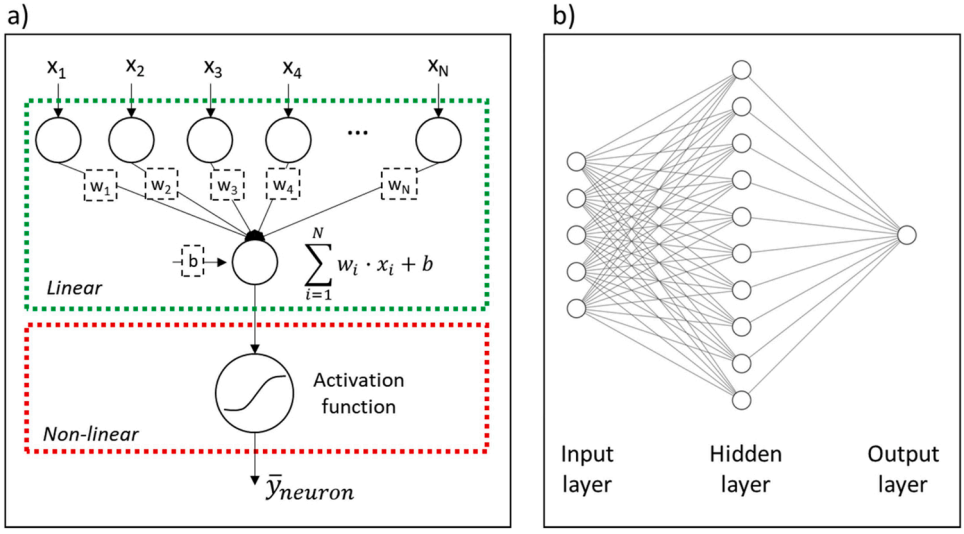

Fig. 2. (a) A neuron structure. It is divided into two parts: first, input information is linearly combined, and then nonlinearity is added to the signal using activation functions in the second stage. (b) Multilayer perceptron structure. Information flows from left to right of the network. It is modified at every layer until it reaches the output layer, where the final prediction is returned.

2.6. Dimensionless correlation application

To benchmark the accuracy of the selected hybrid model, the trained hybrid models were compared to a model obtained with Buckingham Π-theorem [59]. The Π-theorem is widely used in engineering to perform dimensionality reduction over a physical system [60]. Summarising the main statement of the theorem, a problem described by n variables and m primary dimensions can be described using (n-m) dimensionless groups. This methodology is often utilised in chemical engineering to reduce the dimensionality of a physical system and develop a model describing its behaviour using the obtained dimensionless values. In this work, the model was fitted employing the non-linear least-squares method implemented in the *curve_fit* function in SciPy [61] library version 1.6.2. Initially, the model was linearised and fitted over the training set to obtain a more suitable initial guess for the fitting algorithm. The result of this operation was used as the initial guess to fit the non-linear model using the least-squares method. The *dogbox* option was selected as the fitting method of the *curve_fit* function. The confidence interval of the dimensionless model parameters reported Section 3.1 is the standard deviation of the parameter estimated during the fitting. It is calculated using the matrix of covariances returned by the fitting algorithm. More specifically, the diagonal of this matrix represents the estimated variance of the fitted parameters; therefore, the confidence is calculated as the square root of the diagonal of the covariance matrix.

2.7. Definition of the success criteria

The model accuracy and generalisation performances were evaluated using the coefficient of determination (R^2), means squared error (MSE) and median absolute percentage error (MedAPE) both on the train and on the test set. Eqs. (8), (9) and (10) report the R^2, MSE and MedAPE, respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{N_{test}} (y_i - \overline{y_i})^2}{\sum_{i=1}^{N_{test}} (y_i - y_{mean})} \tag{8}$$

$$MSE = \frac{\sum_{i=1}^{N_{test}} (y_i - \overline{y_i})^2}{N_{test}} \tag{9}$$

$$MedAPE = median\left(\left|\frac{y_i - \overline{y_i}}{y_i}\right|\right) \tag{10}$$

In these equations, N_{test} is the number of data points contained in the test set, y_i is the experimental output value of ith point in the test set, $\overline{y_i}$ is the output value predicted at the ith point in the test set, y_{mean} is the average of the output values contained in the test set.

3. Results and discussions

3.1. Buckingham Π-theorem model

A dimensionless regression model was developed by using Buckingham Π-theorem. According to physical information available in the literature, this process can be modelled by using the parameters reported in Eq. (11).

$$K_G a = f(M_w, D_{32}, D_{column}, D_G^{CO_2}, L, G, \rho_G, \rho_L, \mu_G, \mu_L, \sigma_L) \tag{11}$$

where $K_G a$ is overall mass transfer coefficient [mol·m^{-3}·s^{-1}·Pa^{-1}], M_w is molecular weight of the absorbent liquid [kg·mol^{-1}], D_{32} is Sauter mean diameter of droplets [m], D_{column} is diameter of the column [m], $D_G^{CO_2}$ is diffusivity of the CO_2 within gas phase [m^2·s^{-1}], L is liquid flux [kg·m^{-2}·s^{-1}], G is gas flux [kg·m^{-2}·s^{-1}], ρ_G is density of the gas at the inlet of the absorption column [kg·m^{-3}], ρ_L is density of the absorbent liquid at the inlet of the absorption column [kg·m^{-3}], μ_G is viscosity of the gas at the inlet of the absorption column [Pa·s], μ_L is viscosity of the absorbent liquid at the inlet of the absorption column [Pa·s], σ_L is surface tension of the absorbent liquid at the inlet of the absorption column [N·m^{-1}], v_L is velocity of the liquid coming out of the nozzle [m·s^{-1}]. More information about the calculation of these parameters can be found in the Appendix in section A3.

The system is characterised by 12 parameters and 4 fundamental dimensions (kg, m, s, mol). Therefore, according to Buckingham Π-theorem, the system can be described by 8 dimensionless values. We have identified the ones reported in Eq. (12) in this work.

$$\frac{K_G a * M_w}{D_G^{CO_2}} = f(Re_G, Sc_G, Re_L, \frac{L}{G}, \frac{D_{32}}{D_{column}}, \frac{\rho_L}{\rho_G}, \frac{L^2 * D_{32}}{\sigma_L * \rho_L}) \tag{12}$$

where Re_G is the Reynolds number for the gas phase [-], Sc_G is the Schmidt number for the gas phase [-], Re_L is the Reynolds number for the liquid phase [-]. The final dimensionless model has the following structure:

$$\frac{K_G a * M_w}{D_G^{CO_2}} = A * (Re_G)^{\alpha} * (Sc_G)^{\beta} * (Re_L)^{\gamma} * \left(\frac{L}{G}\right)^{\delta} * \left(\frac{D_{32}}{D_{column}}\right)^{\varepsilon} * \left(\frac{\rho_L}{\rho_G}\right)^{\zeta} * \left(\frac{L^2 * D_{32}}{\sigma_L * \rho_L}\right)^{\eta} \tag{13}$$

By performing a parameter fitting on the training data, the following parameters were identified: A = 3.73e11 ± 4.91e-13, α = −7.98 ± 5.94e-01, β = 6.74 ± 9.16e-01, γ = 1.58e-1 ± 6.94e-02, δ = −8.09 ± 5.92e-01, ε = −4.66 ± 3.33e-01, ζ = 9.67 ± 9.22e-01, η = 4.42 ± 2.96e-01. The parameters were estimated using 311 datapoints of the experimental dataset. Further information about the variable boundaries of Eq. (13) can be found in the Appendix in section A3.

Fig. 3 reports quality of the prediction for the dimensionless model. The points with higher error values are located at lower experimental $K_G a$ values ($K_G a$<6). On the other hand, lower error values are associated with the points with higher experimental $K_G a$ values. Here, the presence of the data points is rarer than in the lower experimental output area. Thus, the model returns more accurate predictions in the area where the data points are rarer, while a lower accuracy characterises the area where the data points are more present. This is counterintuitive since the regressive models have higher accuracy where the points are more accumulated and, on the contrary, return poor accuracy where the points are rarer [62]. However, for this case, the optimiser found the minimum of the loss function in a solution that performs more

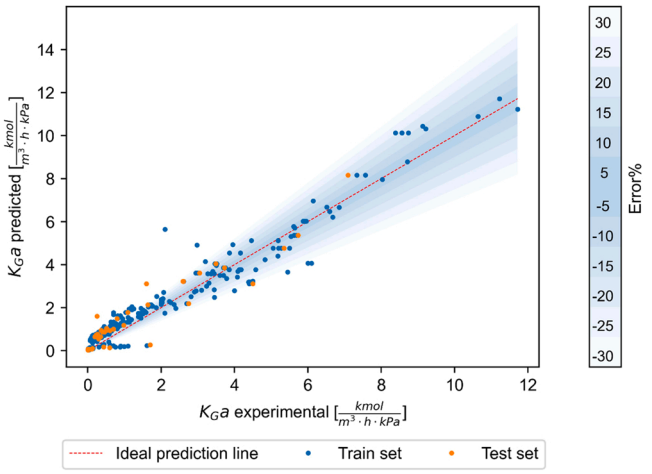

Fig. 3. Prediction performances of the dimensionless model. For low $K_G a$ values, the prediction results are very poor. For high $K_G a$ values, the predictions are accurate on the training and validation set.

accurate predictions to higher K_Ga values. This can be related to the distribution of the experiments in the dataset. Most of the data points utilised in this study were obtained using pressure nozzles and concentration of MEA up to 40% in water (Table1), whereas for the ultrasonic nozzle there were fewer data points all located at high K_Ga values (K_Ga>6). This could increase the clarity of the pattern at higher K_Ga values and decrease it for lower K_Ga values. Thus, for the points located at the lower output area, the optimiser must find an average behaviour among the experiments to reduce the average error. This results in lower prediction accuracy at lower K_Ga values and highly accurate prediction for higher K_Ga values.

The dimensionless model has the lowest prediction accuracy on the train and test sets. Further information can be found in Section 3.3.

3.2. Performance of hybrid modelling

In this section, we compare performances of the hybrid models based on various regressors introduced in the Methods section.

3.2.1. Ridge regression-based hybrid model

The Ridge regression returns poor prediction accuracy both on the training and test sets (Fig. 4 and Table 2). Higher percentage error values are located at lower K_Ga values (K_Ga<6). As previously highlighted (Section 3.1 of this paper), this area is characterised by the output returned by the pressure nozzle. In addition, one can notice poor prediction accuracies of the model at low K_Ga values. For K_Ga< 1, adaptability of the model to the input conditions is very poor. The predictions are often outside the boundaries of ± 30 % errors. Furthermore, the plot shows a vertical line at experimental K_Ga values close to 0. On the contrary, at higher K_Ga values, the Ridge regressor returns predictions with a percentage error value within the range ± 30 %. In this area, the points are generated only by ultrasonic spray. Thus, the hybrid model employing the Ridge regressor returns better predictions for the ultrasonic spray atomisation. Despite the poor prediction capabilities, the Ridge regressor has a non-uniform overfitting behaviour on the training set. As a matter of fact, the R^2 and the MSE metrics are comparable both on the train set and on the test set. On the contrary, the MedAPE results were almost double on the test set compared to the training set. This is related to the points at lower K_Ga values since they have a higher percentage error of the entire K_Ga range (Fig. 4). In other words, the Ridge regression model shows a double behaviour: at low K_Ga values, it overfits the training set, while at higher K_Ga values, it shows good generalisation performances.

Table 2
Performances of the modelling techniques utilised in this work. The decision tree regressor (DTr) is the most performant on the train set. However, the SVMr has the highest generalisation capabilities because of the better performances on the test set.

Model	R^2		MSE		MedAPE	
	Train	Test	Train	Test	Train	Test
Dimensionless	0.94	0.88	0.32	0.38	42%	83%
Ridge	0.95	0.95	0.25	0.20	26%	35%
DTr	0.99	0.99	1.0e-2	2.6e-2	3.3%	11%
SVMr	***0.99***	***0.99***	***1.5e-2***	***2.8e-2***	***6.6%***	***6.4%***
ANN	0.93	0.91	0.36	0.40	23%	33%

The Ridge regressor has similar performances compared to the dimensionless model. Further information can be found in Section 3.3.

3.2.2. Decision tree regressor-based hybrid model

The DTr returns high prediction accuracy obtaining the lower MSE both the train and test sets of this study (Table 2). The percentage errors on the train set remain within the range ± 4 % for the entire output range, and the percentage errors on the test set is ± 11 % on the test set with only a few exceptions (Table 2 and Fig. 5). In addition, the DTr model returns a homogeneous error within the entire output range, and no local increases of the error are identifiable (Fig. 5). In other words, the model does not show any bias with the type of atomisation or MEA concentration. This behaviour of the DTr is related to its capability to create internal sub-models and choose the proper one according to the situation. This is a great advantage compared to all the other regressors investigated in this study. The DTr model uses w_{MEA} as the decision variable for the first layer. It utilises a selection threshold for this variable of 0.56 on the normalised scale. Moreover, the Gini importance index of this variable is 0.62. It is the highest value among the 27 features employed in this model. This means that the major clustering within the data is related to this variable [63]. In physical terms, it means that the concentration of the MEA in the absorbent solution is the most important variable impacting the output value among the model features.

The predictions performed by the DTr model are more accurate than the ones obtained by the dimensionless model. Further information can be found in Section 3.3 of this work.

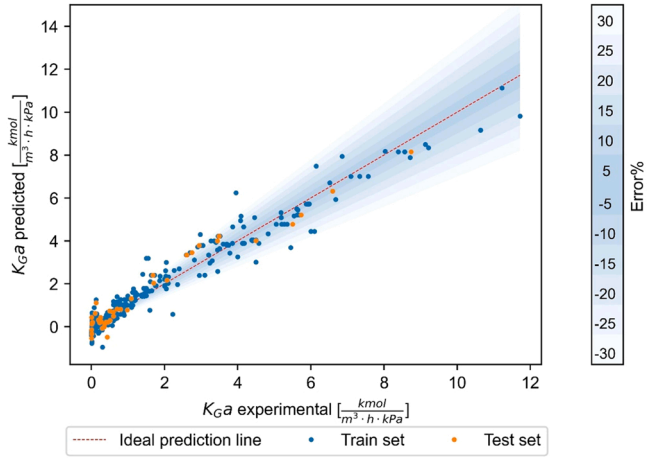

Fig. 4. Prediction performances of the Ridge regression. At low K_Ga values, predictions return a high percentage error. At high K_Ga values, predictions return errors in the range ± 30 %. However, at low K_Ga values, most points are outside these boundaries.

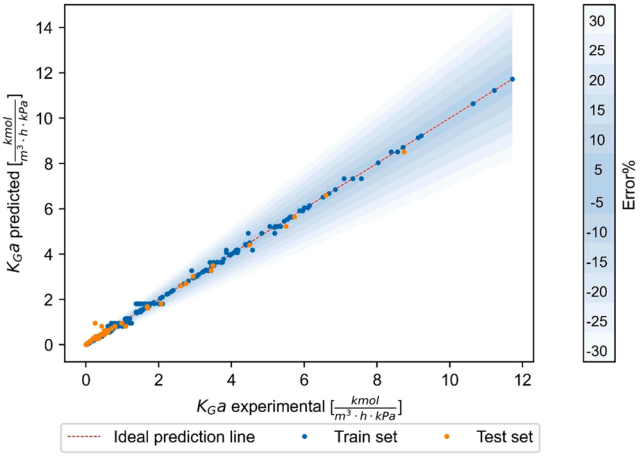

Fig. 5. Prediction performances of the decision tree regression. The prediction quality is uniform within the entire output range without any deviation where the points are rarer. In addition, the quality of the predictions is homogeneous within the entire output range without any deviation. The model shows high generalisation capabilities on the test set.

3.2.3. Support vector machine regressor-based hybrid model

SVMr shows high prediction accuracy and generalisation capabilities (Fig. 6 and Table 2). In Fig. 6, all the points are very close to the ideal prediction line, with only a few exceptions on both sets. As the DTr, also in this case, the prediction accuracy is homogeneous within the entire output range without any bias related to the MEA concentration within the absorbent solution and the spray technique. This modelling technique achieves, together with the DTr, the highest R^2 in this work both on the train set and the test set (see Table 2 in Section 3.3). The high R^2 is related to the low MSE obtained by the SVMr. This translates into a very low MedAPE on the train set and the test set achieved by this modelling technique. Moreover, the SVMr model has a lower loss of accuracy when moving from the train to the test set in the frame of this work (Table 2). In other words, the SVMr has the best generalisation capabilities among the techniques investigated in this study.

The prediction performances of SVMr outstand the performances of the dimensionless model. Further information can be found in Section 3.3 of this work.

3.2.4. Multilayer perceptron regressor-based hybrid model

Fig. 7 reports the prediction capabilities of the hybrid model based on the ANN regressor. The prediction accuracy returned by this modelling technique is poor. This effect is mainly related to the shallow neural network structure. As a matter of fact, it has only one hidden layer containing 116 neurons. The points are highly scattered around the ideal prediction line (Fig. 7). Moreover, the prediction quality is not uniform along the output range. For the K_Ga values located at experimental $K_Ga < 2.5$, prediction points are accumulated around the ideal prediction line with a lower scattering than the higher K_Ga values. From this situation, it is possible to conclude that the MSE for the prediction in this zone is lower compared to values located at higher prediction output. However, because of the small experimental output value, the percentage errors are the highest observed within the entire output range. In fact, some of them go beyond the scale of 30 % set on the graph. This is not observable at high K_Ga values. In the central zone of the plot, identified by the condition of $2.5 < K_Ga < 7.5$, the points are distributed around the ideal prediction line but with a higher scattering grade compared to the lower experimental K_Ga area. This increased the MSE in this zone compared to the output area identified by $K_Ga < 2.5$. Despite that, most of the predictions in this area returned errors in the range $\pm$ 30 % with only a few outliers. Moving along the plot, at experimental $K_Ga > 7.5$ the model predictions accuracy worsens by increasing the output value. The points are always located below the ideal prediction line in this area. This shows how the model has a bias in this area. More specifically, the model outputs always blow the expected value. The bias increases with the output value. This phenomenon is related to the distribution of the data within the dataset. As previously stated, the data distribution is not uniform within the dataset; the data points are more accumulated at low output than at high K_Ga values. This scenario gives the model better accuracy where the data are more accumulated and lower prediction accuracy where the data are rarer [62]. The training of the ANN is affected by this scenario, which lets the model prediction accuracy decrease where the data points are rarer. We suppose this is related to the complexity of the network utilised in this study. In fact, the network has only one hidden layer and should both model the strong nonlinearity of the system and, in the ideal scenario, create a pattern for higher K_Ga values. However, it is not possible to achieve both. Therefore, to obtain a lower loss function, the optimiser found the optimal solution in giving some biases to the network in the higher K_Ga values areas since this decreases the overall training loss function.

The regressor has similar performances compared to the dimensionless model. Further information can be found in Section 3.3.

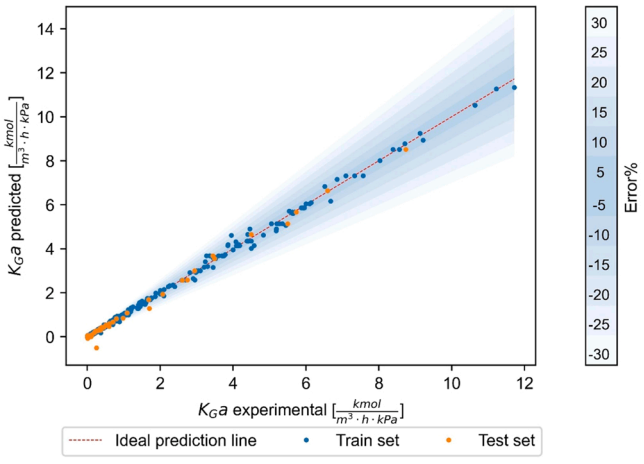

Fig. 6. Prediction performances of the support vector machine regressor both on the train and test sets. The prediction performances are uniform within the entire output range. The model shows high generalisation capabilities on the test set.

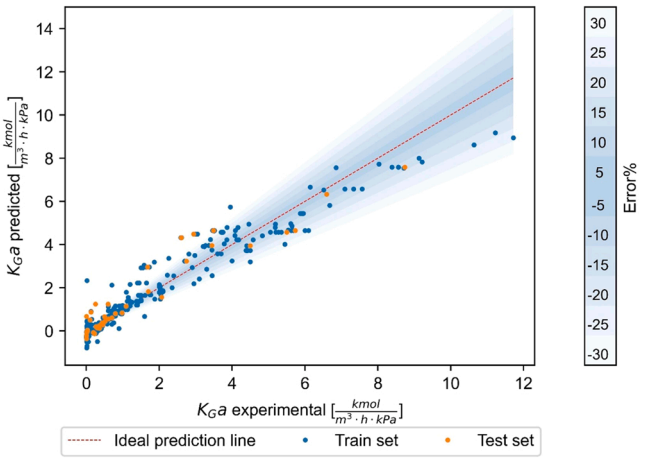

Fig. 7. Prediction performances of the hybrid model obtained with the fully connected artificial neural network. The technique shows poor prediction accuracy both on the training and test sets showing a nonlinear bias within the entire output range.

3.3. Comparison between the models employed in this study

In this section of the paper, the prediction performances of the dimensionless and hybrid models are compared.

3.3.1. Comparison of the hybrid models with the dimensionless model

The dimensionless model has the lowest prediction accuracy on the train and test sets as shown in Table 2. Despite the high value of R^2 on the test set, it is still the lowest experienced in this study. Furthermore, it returns high percentage errors. The MedAPE value is the highest experienced in this work on the train and test set. This is mainly due to the high deviation of the model at low K_Ga values. The low output zone corresponds to processes executed with pressure nozzles and low MEA concentration (Table 1). These two factors could drive the poor prediction capabilities in this area. D_{32} values were calculated using empirical correlations (see Appendix for further information) in this work. D_{32} and aerosol number concentration of the nozzle can be different than the predicted values since these nozzles are not always designed to give a narrow size distribution. There can be deviations from the predicted D_{32} values depending on how well the nozzles are designed. For the ultrasonic nozzle the droplet size distribution was reported by Kayahan et al. [15] and the values had a little deviation from

the calculated D_{32} value. Therefore, the input information has a higher uncertainty on the pressure nozzle than the ultrasonic nozzles. Moreover, low K_Ga are characterised by low MEA concentration. In our previous work we showed that increasing the MEA concentration seems to change the reaction mechanism [15]. The inclusion of this information in the parameters of the dimensionless model is hindered because of the low parameters amount and its model structure. Therefore, the model cannot store this information and during the training prefers to select parameters in such a way that the error at high K_Ga is lower to minimise the cost function.

The Ridge regressor has similar performances compared to the dimensionless model. It returns lower MedAPE compared to the dimensionless model on the train set and slightly better performances on the test set. This is mainly due to the structure of this regressor. The Ridge regression performs a linear fitting over the model parameters, while the dimensionless model requires identifying exponential parameters. Therefore, the loss function of the Ridge regression is less complex than the one of the dimensionless model. This allows the optimiser to perform with the Ridge regressor a better minimisation of the loss function calculated on the training set than the dimensionless model. In addition, the Ridge regression itself can only perform linear correlation. For this reason, the only nonlinearity it can exploit to increase the model accuracy is the one already contained in the input features. In contrast, the parameters of the dimensionless model are exponential. So, identifying the parameters in the dimensionless model adds a layer of nonlinearity that the optimiser can utilise to reduce the mode error.

The predictions performed by the DTr model are more accurate than the ones obtained by the dimensionless model. The R^2 metric is 0.11 points higher than the dimensionless model both on the test sets and 0.05 points higher on the train set. The MSE of the DTr model is one order of magnitude lower than the dimensionless model on the train and the test set. These let the MedAPE of the DTr be one order of magnitude lower on the dimensionless model (Table 2).

The prediction performances of SVMr outstand the performances of the dimensionless model. This is mainly related to the capability of this modelling technique to handle strongly nonlinear behaviour. This effect is related to the possibility of SVMr mapping the model features in a high-dimension space. Here, the optimiser linearly correlates the features map to the output value. This way, the optimiser easily finds the optimal parameters to predict such significant nonlinear behaviour.

Compared to the dimensionless model, the ANN return lower R^2 on the training set and higher MSE metrics showing lower accuracy; however, it also returns lower MedAPE metrics and higher R^2 on the test set than the dimensionless model (Table 2). Overall, it is possible to state that the dimensionless and the ANN models have comparable prediction accuracies both on the train and the test sets. This behaviour is related to the dimension of the ANN employed in this study. The multiple parameters included in the network and the activation function give the possibility to predict nonlinearity; however, the structure is not complex enough to reach the desired level of accuracy. For these reasons, employing an ANN with more layers could significantly improve the prediction accuracy of the model.

3.3.2. Comparison between the hybrid models

The CO_2 capture into MEA is highly non-linear because of the dependence of the reaction kinetics on the MEA concentration and uncertainty related to the surface area (i.e., D_{32} used as model input). In addition, the system behaviour differs within the explored input space, such as liquid flowrate, gas flowrate, their ratio, the MEA to CO_2 ratio etc. This makes the predictions performances of the various hybrid models significantly different (Table 2 and Fig. 4 to Fig. 7).

The hybrid models based on the Ridge regressor and the ANN regressor are the less performant models investigated in this study. It is related to the strong nonlinearity between the inputs and the system output. Despite using monomial and rational permutations of the input variables as model features, the Ridge regressor and the ANN regressor cannot handle such a complex nonlinearity because of their structure. The Ridge regression is a linear regression with an L2 penalty in the loss function; thus, it cannot handle nonlinear systems without the support of pre-processed features. The nonlinearity that a Ridge regression can handle is the same contained in the pre-processed features used as input of the model. Consequently, the pre-processed features were not descriptive enough of the nonlinearity of the system. Therefore, the predictions returned by the hybrid model based on the Ridge regressor have high errors and biases. In principle, the structure of an ANN can handle nonlinear behaviours, especially because of the activation function within the nodes of the network and multiple consecutive layers. However, the optimal network utilised in this study was composed of only one hidden layer employing ReLU activation functions. These two factors significantly decrease the possibility for the structure to handle a complex nonlinear system, such as the case in analysis. For this reason, the network employed in this work returns poor prediction accuracy. They are mainly related to the shallow structure of the employed model. The network has only one hidden layer with 116 nodes. The number of layers was set to one to reduce the number of model parameters. This way, the comparison with the other models employed in this study was fairer. In addition, the training set contained only 311 data points. Therefore, increasing the network structure could increase the data required to have an accurate and generalisable model. Thus, employing a network with more hidden layers could improve the accuracy and generalisation capabilities of the network; however, this could require more training data points.

The most performant hybrid models in this study employ the DTr and the SVMr. These two structures have individual peculiarities allowing them to handle strongly nonlinear and complex systems, such as the case reported in this study. The DTr creates internal sub-models and characterises each of them to a specific area of the system. This significantly increases the capability of the model to expand the case-specificness according to the situation reported by the input variables of the model. On the other hand, the SVMr has the peculiarity of creating support vectors within a higher dimensional space than the one reported in the input variables. The optimal space is identified when the support vectors and the desired output can be correlated using a hyperplane (i.e., a linear correlation). In such a way, the SVMr handles complex nonlinear behaviours using a limited number of parameters. This also decreases the opportunity for the model to overfit the training data.

The better prediction capabilities of the SVMr and DTr can be explained with the different behaviour of the system when employing different spray technologies and MEA concentration. The spray technology impacts the droplet size distribution. In addition, MEA concentration affects the absorption reaction path. The SVMr and the DTr have modelling structures suitable for classification, therefore their structure is more prone to create decision path than the other models (i.e., Ridge regression, ANN and dimensionless model) that store the information in parameters used also for the prediction.

The DTr returns a higher accuracy on the train set. However, higher generalisation capabilities are associated with the SVMr (Table 2). The MSE and MedAPE of the DTr hybrid model significantly increase when moving from the train to the test set. In contrast, the same metrics of the SVMr hybrid model do not experience such a significant increase. The higher overfitting phenomena of the DTr compared to the SVMr are related to the possibility of further creating ad-hoc sub-models to learn the pattern contained in the train set. However, the sub-models have poor generalisation capabilities because of an over-refinement on the training data. Even if in this study a cross-validation approach was applied to refine the hyper-parameter of the various regressor, this is intrinsically present in the DTr technique. In contrast, this problem is not present in the SVMr since it uses a parametric modelling approach, and the possibility of overfitting the training data is lower.

Because of the reduced number of parameters, the lower overfitting phenomena, and the outstanding prediction accuracies, we suggest

employing SVMr to model absorption columns for future works. They offer remarkable generalisation among column sizes, spraying techniques and absorbent liquid formulations. This modelling approach can be used for predictive modelling and control optimisation. On the other hand, DTr can be used to investigate the impact of each feature on the final output and how the sub-models are clustered but strongly discourage using this regression technique to optimise operational optimisation.

DTr and SVMr return higher accuracy than the dimensionless model obtained with the Buckingham Π-theorem. This behaviour is associated with features extraction capabilities. In fact, it is well known how the CO_2 absorption into MEA is driven by reactive absorption [64]; however, explicit information about the reaction between MEA and CO_2 was not included in the training data of both modelling approaches. The absence of this information also drives the lack of accuracy in the dimensionless model. Despite this, the SVMr and the DTr could extract and understand this internal correlation between the variables to obtain higher prediction capability.

Using the monomial and rational interactions as input increased the model prediction capabilities. After the monomial and rational interactions computation, a statistical evaluation of each input variable dependency (interactive or not) over the final output is performed. The adopted procedure computes the p-values of the interaction with the K_Ga value. Following, the independent variables with p-values$<$ 0.05 are selected. This procedure returns a set of 27 variables, and around 80 % (i.e., 21 variables) contain interactions. The high presence of interactive variables indicates that, with a given model structure, the prediction performances must be better when adding the interaction parameters since the model has input variables strongly correlated to the final output value. In addition, the highly correlated input variables are not interdependent because of the selection performed with the Pearson test (further information can be found in Section 2.3.1 of this work). Assume the computation and selection of the interaction variables are not performed. In that case, the model must extract these features, which is hindered by the simple model structures used in our work.

Overall, the most performant hybrid models employed in this study were the one including DTr or SVMr. As previously stated, despite the better performance of the DTr model on the train set compared to the SVMr, the latter showed improved generalisation capabilities on the test set. For these reasons, the usage of the DTr is strongly discouraged when predicting the K_Ga of the system. The employment of SVMr is suggested for such a task.

3.3.3. Extension of the model to packed columns and other solvents

The models proposed in this paper were characterised for spray columns. One of the main bottlenecks in the spray columns is the evaluation of the surface area of the spray. Not many models are available in the literature to predict such a value, and often it must be measured experimentally. For all these reasons, the hybrid models employed in this work did not include the surface area of the spray in the input variable space. In case the methodology proposed in this work has to be extended to packed columns, the expected surface area of the liquid could be included in the model input because of the large availability of correlation to predict the wet surface of a packed bad for absorption [9]. In alternative to the wet surface of the packed bad, the overall surface area of the packing and some geometric characteristics of the packing (e.g., void degree, geometrical dimensions, and wettability of the packing material) can be used as model input. In both cases, the information about the wet or overall packing area must replace the Sauter mean diameter value in the model input.

The current model was trained for aqueous MEA mixtures. In case the methodology has to be applied to model the capture using other chemicals, the model structure proposed in this paper is already suitable. Therefore, if the structure has to be applied to a new absorbent liquid, the physical properties of such a liquid (i.e., density, surface tension and viscosity) must be calculated using validated models describing them. However, the proposed structure is not flexible to different formulations of absorbent liquid (e.g., to make predictions on a mixture containing MEA and water, and on a mixture containing diethanolamine and water with the same trained model). In such a case, the model input can be expanded by adding chemical information about the molecules involved in the absorption or information about the reaction kinetics (e.g., the Arrhenius constant of the main reactions driving the absorption process).

4. Conclusions

In this study, we analysed the performances of various machine learning techniques to create a hybrid model of the mass-transfer coefficient (K_Ga) of the CO_2 capture process executed with aqueous and pure monoethanolamine (MEA) solutions in spray columns. This work aimed to analyse various machine learning techniques to identify the most suitable one for this purpose. The data to train the hybrid models were obtained from the literature. The results were compared with the state of the art in K_Ga modelling represented by the dimensionless model obtained using the Buckingham Π-theorem. The hybrid models developed in this study were based on the Ridge regression, decision tree regressor (DTr), support vector machine regressor (SVMr) and artificial neural network (ANN). The accuracy and the generalisation capabilities of the model were investigated. This study identified that the most proficient models for prediction accuracy were the SVMr and the DTr. These models were initially designed for classification tasks and readapted to regression. As a result, they have peculiar capabilities to handle nonlinear behaviours showing the median of the absolute percentage error (MedAPE) on the train set to be 6.6 % and 3.3 % respectively. The SVMr reported the highest accuracy and generalisation capabilities among different column sizes, spray technologies and absorbent liquid formulations in the frame of this work. The value of the MedAPE obtained on the test set was 6.4 % for the SVMr and 11% for the DTr. For this reason, we suggest employing the SVMr for the predictions of the K_Ga for CO_2 capture in spray columns. On the other hand, the DTr results are even more accurate than the SVMr on the train set, however, when the trained model is evaluated on the test set, the accuracy drops, and the median value of the percentage error increases significantly. Because of the poor generalisation capabilities of the DTr, we strongly discourage using this modelling technique to predict the K_Ga of the system. However, it is still a powerful tool for evaluating the importance of each variable in the computation of the output. It can thus still be a support tool for optimising the spray columns for CO_2 capture. The less accurate models encountered in this study were the Ridge and ANN regressions. The Ridge regressor showed higher accuracies than the dimensionless model while the ANN has comparable performances than the dimensionless model. The poor prediction capabilities of the ANN are mainly related to its shallow structure. However, they had lower prediction accuracies and generalisation capabilities than SVMr and DTr. All the aforementioned hybrid models were also compared with the dimensionless model obtained using Buckingham Π-theorem. All of them return better prediction capabilities because of their complex structures. Further development of this study is the investigation of more complex neural networks to compute the K_Ga of the system and the investigation of the modelling performance where other amines than MEA are included in the formulation of the absorbent liquid.

CRediT authorship contribution statement

Ulderico Di Caprio: Methodology, Investigation, Data curation, Conceptualisation, Formal analysis, Visualisation, Software, Validation, Writing - original draught, Writing - review & editing. **Min Wu**: Writing - review & editing. **Florence Vermeire**: Writing - review & editing. **Tom Van Gerven**: Writing - review & editing, Supervision. **Peter Hellinckx**: Writing - review & editing, Supervision. **Steffen Waldherr**: Writing - review & editing, Supervision. **Emine Kayahan**: Data curation,

Conceptualisation, Writing - review & editing, Supervision. **M. Enis Leblebici**: Funding acquisition, Resources, Project administration, Supervision, Writing - review & editing.

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.

Acknowledgements

The authors acknowledge funding from VLAIO-Catalisti Moonshot projects “Intensification of CO2 Capture Processes” (HBC.2019.0109 – CAPTIN and HBC.2021.0254 – CAPTIN2) and “Real-time data-assisted process development and production in chemical applications” (HBC.2020.2455 – DAP2CHEM). The authors declare no competing interests.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jcou.2023.102452.

References

[1] R.P. Allan, P.A. Arias, S. Berger, J.G. Cassou, C. Cassou, D. Chen, A. Cherchi, et al., IPCC, 2021: Summary for Policymakers, 2021.
[2] A.F. Ghoniem, Needs, resources and climate change: clean and efficient conversion technologies, Prog. Energy Combust. Sci. 37 (2011) 15–51, https://doi.org/10.1016/J.PECS.2010.02.006.
[3] P. Luis, Use of monoethanolamine (MEA) for CO_2 capture in a global scenario: consequences and alternatives, Desalination 380 (2016) 93–99, https://doi.org/10.1016/J.DESAL.2015.08.004.
[4] Y. Wang, L. Zhao, A. Otto, M. Robinius, D. Stolten, A review of post-combustion CO_2 capture technologies from coal-fired power plants, Energy Procedia 114 (2017) 650–665, https://doi.org/10.1016/J.EGYPRO.2017.03.1209.
[5] L.C. Law, N. Yusoff Azudin, S.R. Abd. Shukor, Optimization and economic analysis of amine-based acid gas capture unit using monoethanolamine/methyl diethanolamine, Clean. Technol. Environ. Policy 20 (2018) 451–461, https://doi.org/10.1007/s10098-017-1430-1.
[6] C. Kale, A. Górak, H. Schoenmakers, Modelling of the reactive absorption of CO_2 using mono-ethanolamine, Int. J. Greenh. Gas Control (2013), https://doi.org/10.1016/j.ijggc.2013.05.019.
[7] M. Afkhamipour, M. Mofarahi, Review on the mass transfer performance of CO_2 absorption by amine-based solvents in low- and high-pressure absorption packed columns, RSC Adv. 7 (2017) 17857–17872, https://doi.org/10.1039/c7ra01352c.
[8] G.Q. Wang, X.G. Yuan, K.T. Yu, Review of mass-transfer correlations for packed columns, Ind. Eng. Chem. Res. 44 (2005) 8715–8729, https://doi.org/10.1021/ie050017w.
[9] D. Flagiello, A. Parisi, A. Lancia, F. di Natale, A review on gas-liquid mass transfer coefficients in packed-bed columns, ChemEngineering 5 (2021), https://doi.org/10.3390/chemengineering5030043.
[10] N. Razi, H.F. Svendsen, O. Bolland, Assessment of mass transfer correlations in rate-based modeling of a large-scale CO_2 capture with MEA, Int. J. Greenh. Gas Control 26 (2014) 93–108, https://doi.org/10.1016/J.IJGGC.2014.04.019.
[11] D.J. Heldebrant, P.K. Koech, V.A. Glezakou, R. Rousseau, D. Malhotra, D.C. Cantu, Water-lean solvents for post-combustion CO_2 capture: fundamentals, uncertainties, opportunities, and outlook, Chem. Rev. 117 (2017) 9594–9624, https://doi.org/10.1021/acs.chemrev.6b00768.
[12] R.R. Wanderley, H.K. Knuutila, Mapping diluents for water-lean solvents: a parametric study, Ind. Eng. Chem. Res. 59 (2020) 11656–11680, https://doi.org/10.1021/acs.iecr.0c00940.
[13] J. Kuntz, A. Aroonwilas, Performance of spray column for CO_2 capture application, Ind. Eng. Chem. Res. 47 (2008) 145–153, https://doi.org/10.1021/ie061702l.
[14] X.M. Wu, Z. Qin, Y.S. Yu, Z.X. Zhang, Experimental and numerical study on CO_2 absorption mass transfer enhancement for a diameter-varying spray tower, Appl. Energy 225 (2018) 367–379, https://doi.org/10.1016/j.apenergy.2018.04.053.
[15] E. Kayahan, U. di Caprio, A. van den Bogaert, M.N. Khan, M. Bulut, L. Braeken, T. van Gerven, M.E. Leblebici, A new look to the old solvent: mass transfer performance and mechanism of CO_2 absorption into pure monoethanolamine in a spray column, Chem. Eng. Process. - Process. Intensif. 184 (2023), 109285, https://doi.org/10.1016/j.cep.2023.109285.
[16] T.N. Borhani, M. Wang, Role of solvents in CO_2 capture processes: the review of selection and design methods, Renew. Sustain. Energy Rev. 114 (2019), 109299, https://doi.org/10.1016/j.rser.2019.109299.
[17] U. di Caprio, E. Kayahan, M. Wu, S. Mercelis, P. Hellinckx, T. van Gerven, S. Waldherr, M.E. Leblebici, Optimization of an artificial neural network structure for modelling carbon capture in spray columns, Comput. Aid. Chem. Eng. 51 (2022) 1411–1416, https://doi.org/10.1016/B978-0-323-95879-0.50236-8.
[18] B. Rajmohan, S. Biswas, B.C. Meikap, Removal of SO2 from industrial effluents by a novel twin fluid air-assist atomized spray scrubber, Ind. Eng. Chem. Res 47 (2008) 7833–7840, https://doi.org/10.1021/ie900188r.
[19] T. Tanda, K. Shirai, Y. Matsumura, H. Kitahara, New correlation for mass transfer characteristics of spray column, Ind. Eng. Chem. Res 50 (2011) 13554–13560, https://doi.org/10.1021/ie200989k.
[20] M.R. Dobbelaere, P.P. Plehiers, R. van de Vijver, C.V. Stevens, K.M. van Geem, Machine learning in chemical engineering: strengths, weaknesses, opportunities, and threats, Engineering 7 (2021) 1201–1211, https://doi.org/10.1016/J.ENG.2021.03.019.
[21] L. Chiang, B. Lu, I. Castillo, Big data analytics in chemical engineering, Annu Rev. Chem. Biomol. Eng. 8 (2017) 63–85, https://doi.org/10.1146/annurev-chembioeng-060816-101555.
[22] N.A. Menad, A. Hemmati-Sarapardeh, A. Varamesh, S. Shamshirband, Predicting solubility of CO_2 in brine by advanced machine learning systems: application to carbon capture and sequestration, J. CO_2 Util. 33 (2019) 83–95, https://doi.org/10.1016/J.JCOU.2019.05.009.
[23] V. Venkatraman, B.K. Alsberg, Predicting CO_2 capture of ionic liquids using machine learning, J. CO_2 Util. 21 (2017) 162–168, https://doi.org/10.1016/j.jcou.2017.06.012.
[24] M. Mesbah, S. Shahsavari, E. Soroush, N. Rahaei, M. Rezakazemi, Accurate prediction of miscibility of CO_2 and supercritical CO_2 in ionic liquids using machine learning, J. CO_2 Util. 25 (2018) 99–107, https://doi.org/10.1016/j.jcou.2018.03.004.
[25] Z. Zhang, H. Li, H. Chang, Z. Pan, X. Luo, Machine learning predictive framework for CO_2 thermodynamic properties in solution, J. CO_2 Util. 26 (2018) 152–159, https://doi.org/10.1016/j.jcou.2018.04.025.
[26] C. Li, H. Liu, M. Xiao, X. Luo, H. Gao, Z. Liang, Thermodynamics and ANN models for predication of the equilibrium CO_2 solubility in aqueous 3-dimethylamino-1-propanol solution, Int. J. Greenh. Gas. Control 63 (2017) 77–85, https://doi.org/10.1016/j.ijggc.2017.05.007.
[27] H. Saghafi, M. Arabloo, Modeling of CO_2 solubility in MEA, DEA, TEA, and MDEA aqueous solutions using AdaBoost-decision tree and artificial neural network, Int. J. Greenh. Gas Control 58 (2017) 256–265, https://doi.org/10.1016/j.ijggc.2016.12.014.
[28] M.M. Ghiasi, S. Abedi-Farizhendi, A.H. Mohammadi, Modeling equilibrium systems of amine-based CO_2 capture by implementing machine learning approaches, Environ. Prog. Sustain. Energy 38 (2019), 13160, https://doi.org/10.1002/ep.13160.
[29] K. Fu, G. Chen, Z. Liang, T. Sema, R. Idem, P. Tontiwachwuthikul, Analysis of mass transfer performance of monoethanolamine-based CO_2 absorption in a packed column using artificial neural networks, Ind. Eng. Chem. Res. 53 (2014) 4413–4423, https://doi.org/10.1021/ie403259g.
[30] K. Onda, H. Takeuchi, Y. Okumoto, Mass transfer coefficients between gas and liquid phases in packed columns, J. Chem. Eng. Jpn. 1 (1968) 56–62, https://doi.org/10.1252/jcej.1.56.
[31] V.Y. Valera, M.C. Codolo, T.D. Martins, Artificial neural network for prediction of SO2 removal and volumetric mass transfer coefficient in spray tower, Chem. Eng. Res. Des. 170 (2021) 1–12, https://doi.org/10.1016/J.CHERD.2021.03.008.
[32] M. von Stosch, R. Oliveira, J. Peres, S. Feyo de Azevedo, Hybrid semi-parametric modeling in process systems engineering: past, present and future, Comput. Chem. Eng. 60 (2014) 86–101, https://doi.org/10.1016/j.compchemeng.2013.08.008.
[33] Y. Tamhankar, B. King, J. Whiteley, T. Cai, K. McCarley, M. Resetarits, C. Aichele, Spray absorption of CO_2 into monoethanolamine: mass transfer coefficients, dropsize, and planar surface area, Chem. Eng. Res. Des. 104 (2015) 376–389, https://doi.org/10.1016/j.cherd.2015.08.012.
[34] S. Zimmermann, M.O. Schmid, B. Klein, G. Scheffknecht, Experimental studies on spray absorption with the post combustion CO_2 capture pilot-plant CASPAR, Energy Procedia 114 (2017) 1325–1333, https://doi.org/10.1016/J.EGYPRO.2017.03.1252.
[35] O. Seyboth, S. Zimmermann, B. Heidel, G. Scheffknecht, Development of a spray scrubbing process for post combustion CO_2 capture with amine based solvents, Energy Procedia 63 (2014) 1667–1677, https://doi.org/10.1016/J.EGYPRO.2014.11.176.
[36] J. Benesty, J. Chen, Y. Huang, I. Cohen, Optimal filters in the time domain, in: Noise Reduction in Speech Processing, Springer, 2009, pp. 1–18, https://doi.org/10.1007/978-3-642-00296-0_7.
[37] D. Berrar, Cross-validation, Encycl. Bioinform. Comput. Biol. (2019) 542–545, https://doi.org/10.1016/B978-0-12-809633-8.20349-X.
[38] T. Fushiki, Estimation of prediction error by using K-fold cross-validation, Stat. Comput. 21 (2011) 137–146, https://doi.org/10.1007/s11222-009-9153-8.
[39] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Ron1995 (Ed.), Proceedings of the 14th International Joint Conference on Artificial Intelligence, Vol. 2, Morgan Kaufmann Publishers Inc, San Francisco, CA, USA, 1995, pp. 1137–1143, https://doi.org/10.5555/1643031.1643047.
[40] C. Ambroise, G.J. McLachlan, Selection bias in gene extraction on the basis of microarray gene-expression data, Proc. Natl. Acad. Sci. USA 99 (2002) 6562–6566, https://doi.org/10.1073/pnas.102102699.

[41] H. Yarveicy, H. Saghafi, M.M. Ghiasi, A.H. Mohammadi, Decision tree-based modeling of CO 2 equilibrium absorption in different aqueous solutions of absorbents, Environ. Prog. Sustain. Energy 38 (2019) S441–S448, https://doi.org/10.1002/ep.13128.
[42] O. Ivanciuc, Applications of support vector machines in chemistry, Rev. Comput. Chem. (2007) 291–400, https://doi.org/10.1002/9780470116449.ch6.
[43] A.E. Horel, Applications of ridge analysis to regression problems, Chem. Eng. Prog. 58 (1962) 54–59.
[44] D.M. Himmelblau, Applications of artificial neural networks in chemical engineering, Korean J. Chem. Eng. 17 (2000) 373–392, https://doi.org/10.1007/BF02706848.
[45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830. ⟨http://jmlr.org/papers/v12/pedregosa11a.html⟩.
[46] A.E. Hoerl, R.W. Kennard, Ridge regression: applications to nonorthogonal problems, Technometrics 12 (1970) 69–82, https://doi.org/10.1080/00401706.1970.10488635.
[47] M. Somvanshi, P. Chavan, S. Tambade, S.V. Shinde, A review of machine learning techniques using decision tree and support vector machine, In: Proceedings of the 2016 International Conference on Computing Communication Control and Automation (ICCUBEA), 2016: pp. 1–7. https://doi.org/10.1109/ICCUBEA.2016.7860040.
[48] L. Li, X. Zhang, Study of data mining algorithm based on decision tree, In: Proceedings of the 2010 International Conference On Computer Design and Applications, 2010: pp. V1–155-V1–158. https://doi.org/10.1109/ICCDA.2010.5541172.
[49] Rahul, A. Gupta, A. Bansal, K.Roy, Solar energy prediction using decision tree regressor, In: Proceedings of the 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021: pp. 489–495. https://doi.org/10.1109/ICICCS51141.2021.9432322.
[50] H. Bhavsar, M.H. Panchal, A review on support vector machine for data classification, In: Proceedings of the International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Citeseer, 2012: pp. 185–189. http://text2fa.ir/wp-content/uploads/Text2fa.ir-A-Review-on-Support-Vector-Machine.pdf.
[51] G. McDonald, N. García-Pedrajas, C. Macdonald, I. Ounis, A Study of SVM kernel functions for sensitivity classification ensembles with POS sequences, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Association for Computing Machinery, New York, NY, USA, 2017, pp. 1097–1100, https://doi.org/10.1145/3077136.3080731.
[52] M. Awad, R. Khanna, Support vector regression, in: Efficient Learning Machines, Apress, Berkeley, CA, 2015, pp. 67–80, https://doi.org/10.1007/978-1-4302-5990-9_4.
[53] A. Zendehboudi, M.A. Baseer, R. Saidur, Application of support vector machine models for forecasting solar and wind energy resources: a review, J. Clean. Prod. 199 (2018) 272–285, https://doi.org/10.1016/j.jclepro.2018.07.164.
[54] S. Shanmuganathan, Artificial neural network modelling: an introduction, in: S. Shanmuganathan, S. Samarasinghe (Eds.), Artificial Neural Network Modelling, Springer International Publishing, Cham, 2016, pp. 1–14, https://doi.org/10.1007/978-3-319-28495-8_1.
[55] O.I. Abiodun, A. Jantan, A.E. Omolara, K.V. Dada, N.A.E. Mohamed, H. Arshad, State-of-the-art in artificial neural network applications: a survey, in: Heliyon, 4, 2018, e00938, https://doi.org/10.1016/J.HELIYON.2018.E00938.
[56] J. Panerati, M.A. Schnellmann, C. Patience, G. Beltrame, G.S. Patience, Experimental methods in chemical engineering: artificial neural networks–ANNs, Can. J. Chem. Eng. 97 (2019) 2372–2382, https://doi.org/10.1002/cjce.23507.
[57] R. Hecht-Nielsen, Theory of the backpropagation neural network, Neural Netw. Percept. (1992) 65–93, https://doi.org/10.1016/B978-0-12-741252-8.50010-8.
[58] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, ArXiv (2014), https://doi.org/10.48550/arXiv.1412.6980.
[59] E. Buckingham, On physically similar systems; illustrations of the use of dimensional equations, Phys. Rev. 4 (1914) 345–376, https://doi.org/10.1103/PhysRev.4.345.
[60] A.A. Sonin, A generalization of the ∏-theorem and dimensional analysis, Proc. Natl. Acad. Sci. USA 101 (2004) 8525–8526, https://doi.org/10.1073/pnas.0402931101.
[61] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Y. Ilhan Polat, E.W. Feng, J. Moore, D. VanderPlas, J. Laxalde, R. Perktold, I. Cimrman, E.A. Henriksen, C.R. Quintero, A.M. Harris, A. H. Archibald, F. Ribeiro, P. Pedregosa, van Mulbregt, SciPy 1.0 Contributors, {SciPy} 1.0: fundamental algorithms for scientific computing in python, Nat. Methods 17 (2020) 261–272, https://doi.org/10.1038/s41592-019-0686-2.
[62] R.P. Ribeiro, N. Moniz, Imbalanced regression and extreme value prediction, Mach. Learn 109 (2020) 1803–1835, https://doi.org/10.1007/s10994-020-05900-9.
[63] S.Sivagama Sundhari, A knowledge discovery using decision tree by Gini coefficient. Proceedings of the 2011 International Conference on Business, Engineering and Industrial Applications, IEEE, 2011, pp. 232–235, https://doi.org/10.1109/ICBEIA.2011.5994250.
[64] H.-B. Xie, Y. Zhou, Y. Zhang, J.K. Johnson, Reaction mechanism of monoethanolamine with CO_2 in aqueous solution from molecular modeling, J. Phys. Chem. A 114 (2010) 11844–11852, https://doi.org/10.1021/jp107516k.