An Efficient Implementation of Artificial Neural Networks with K-fold Cross-validation for Process Optimization

Kathiravan Srinivasan[1], Aswani Kumar Cherukuri[1], Durai Raj Vincent[1], Ashish Garg[1], Bor-Yann Chen[2]

[1] School of Information Technology and Engineering, Vellore Institute of Technology, India
[2] Department of Chemical and Materials Engineering, National Ilan University, Taiwan
kathiravan.srinivasan@vit.ac.in, aswani@vit.ac.in, pmvincent@vit.ac.in, ashish.garg2014@vit.ac.in, bychen@niu.edu.tw

Abstract

A novel approach for building an Artificial Neural Network (ANN) to reconstruct experimental design data using Levenberg-Marquardt optimizer and validating obtained model based on K-fold cross-validation is implemented. In this approach it matters less how the data gets divided, every data point gets to be in test set precisely once and gets to be in a training set (k-1) times. Further, this helps to exclude overfitting of the model on training data and better predictions over unseen data. Moreover, this is the most significant strength and advantage of this approach. Also, by using this approach for validation and Levenberg-Marquardt optimizer two model layers, namely, the input layer consisting of multivariable nodes and the output layer consisting output node, will be conceptually created and validated. The internal layers will be incorporated based on the complexity of the problem. The performance of the obtained model is evaluated using the coefficient of determination. Besides, it is found to have excellent correspondence with experimental results. The method is also compared with existing methods based on model validation, and it shows the much-improved capability to predict optimal results.

Keywords: Artificial Neural Network, K-fold cross-validation, Response surface methodology, Levenberg-Marquardt optimizer

1 Introduction

As known, statistical experimental design with Response Surface Methodology (RSM) is popularly used in engineering for optimization. RSM was first mentioned by Box and Wilson [1] as a prediction tool for experimental design in engineering applications. However, RSM still has certain limitations as it needs a proper experimental design (e.g., well-distributed and non-biased factorial design) for modeling. Moreover, it could not possess learning capability via present experimental data. RSM models are usually accurate only for test range of input parameters (e.g., interpolation, but not extrapolation) Alternatively, the development of higher order of RSM models requires a significant number of experiments to be performed with sufficient calculation and computation; this also limits the use of RSM for highly non-linear systems.

In consequence, to fit high non-linearity and to have better predictions by learning over data; Machine learning (ML) [2] is first used to have a wide range of applications. It provided the construction of algorithms to learn from and make predictions over databases. It is already employed in many computing tasks such as (i) ML web: A toolkit for machine learning on the web, which enables machine learning for web applications [3], (ii) Predicting dissolution kinetics of silicate glasses using machine learning [2], (iii) Machine Learning in medical imaging [4] for image-based diagnosis, disease prognosis and risk assessment, (iv) Machine learning for Gravity Spy: Glitch classification and dataset [5] that combines crowdsourcing with machine learning to help categorize the glitches v) A machine learning scheme to synchronization of automata [6], to predict reset word length for automata. Furthermore, there are several approaches in the field of machine learning to train the data such as Decision tree learning [7], Association rule learning [8], Artificial Neural Networks [9], Deep Learning [10], Inductive logic programming [11], Support Vector Machines [12], Clustering [13], Bayesian Networks [14], Reinforcement learning [15], Representation learning [16], Similarity and metric learning [17], Sparse Dictionary learning [18]. Among them, the Artificial Neural Network (ANN) is adopted herein for study as it can converge on a small set of data via Levenberg-Marquardt [19] learning optimizer.

Also, Artificial Neural Network is the information processing chemo-metric technique created to model non-linear transformation, which simulates essential learning mechanisms of the human brain. Moreover, due to its ability to fit non-linear behavior, it is widely applied in many fields like (i) in Engineering for estimating Biogas Engine Performance [20] (ii) in

*Corresponding Author: Bor-Yann Chen; E-mail: bychen@niu.edu.tw
DOI: 10.3966/160792642019072004020

medical as a SpO_2 measurement for clinical Management Systems [21] (iii) in Energy to predict the performance of Solar Collector Systems [22] (iv) in agriculture as modeling of extraction kinetics of essential oil from tarragon using ultrasound pre-treatment [23] (v) in computational science as modeling of nanostructured memristor device characteristics [24].

Regarding practical engineering applications, there are some papers mentioned technical feasibility of ANN and RSM interactive modeling on different processes for system optimization. For instance, Kartic et al. [25] studied the removal of high concentration of sulfate from the pigment industry using ANN and RSM modeling approaches. Sabour and Amiri [26] studied ANN and RSM modeling of simultaneous optimization of multiple targets in Fenton treatment of landfill leachate; Ohale et al. [27] showed Optimal factor evaluation for the dissolution of alumina from Azaraegbelu clay in acid solution using RSM and ANN modeling.

Even these were proposed for use, data validation from model prediction remained open for further assessment to practicability. Among different validation approaches like Resubstitution, Random subsampling, holdout cross-validation approach is used to compare and validate models in studies mentioned above. The advantage of this approach is that it takes less time for computation, but its result can have colossal inconsistency. The evaluation may entirely depend on which experimental data points goes in training part and which goes in test part. Thus, the evaluation will significantly depend on how the division of train and test part is made.

Regarding the improvement of model validation, K-fold cross-validation [28] is a better approach over the holdout method for validation on experimental design data. The dataset is divided into k sub-parts, and the holdout method is iterated k times. Each time, out of k subparts of data one part is used for testing and the other (k-1) subparts go for training. The advantage of this method is every experimental design data point gets to be in a test subpart and train subpart; this helps to exclude overfitting of the model on training data. The shortcoming of this approach is that the model needs to iterate k times which involves much computation. Hence, more investigation is needed to build an ANN model on experimental design data using Levenberg-Marquardt optimizer and validating using K-fold cross-validation. Here, K is ranged from 5-10 to compare performance with RSM models for correct evaluation of which model is better to predict the optimal experimental output(s)/condition(s).

The significance of this first-attempt study is to show the promising predictive capability of an ANN over experimental design data while validating models using K-fold cross-validation. Moreover, propelled from biological sensory systems and mind structure, Artificial Neural Networks (ANN) could be viewed as data handling frameworks, which permit elaboration of numerous unique methods covering an expansive field of utilization for the proposed approach. Among their most engaging properties, one can cite their learning and speculation abilities. In the event that an expansive number of works are concerned with the theoretical and practical parts of the ANN, just a couple are accessible concerning their true modern application abilities to achieve better results. The RSM entitles to defining and forecasting the relationship prevailing between the response and explanatory variables. Enabling to use a continuous set of experiments leading to an optimal response. These models assist to efficiently determine and apply the actions depending on approximation though only a few information is available for the process. RSM along with ANN enables process optimization depicting the response variable which is influenced by various independence variables. It is found from simulations on 3 set of experimental design data that ANN has good capability in learning and making predictions on experimental design data. Also, system performance is also compared with RSM, and it is found that the ANN owns much better capability to predict experimental design data and to optimize the maximum efficiency of production.

2 Materials and Methods

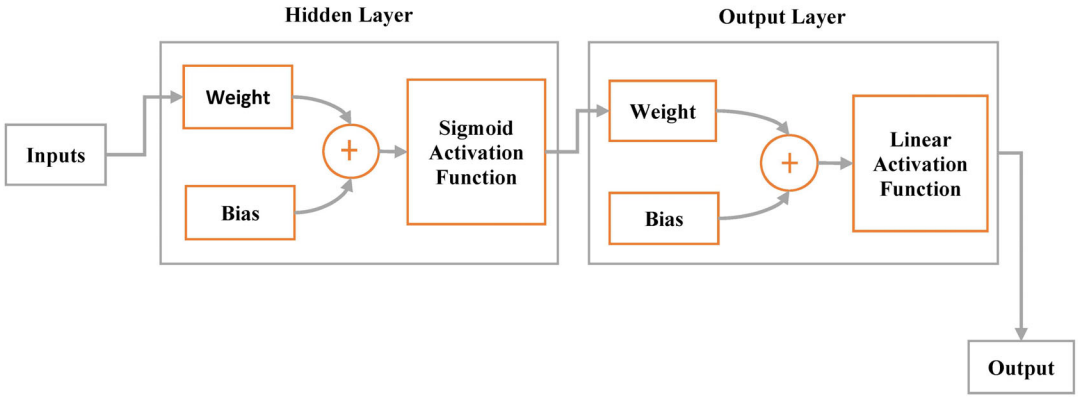
2.1 Adopted Database

Experimental design data developed by Hamid et al [29] for predicting capability of carboxylated cellulose nanowhiskers for the remediation of copper from water, Ohale et al [27] for dissolution of alumina from Azaraegbelu clay and Khatti et al [30] for modeling electrospinning process of polycaprolactone to predict fiber diameter are used in this study.

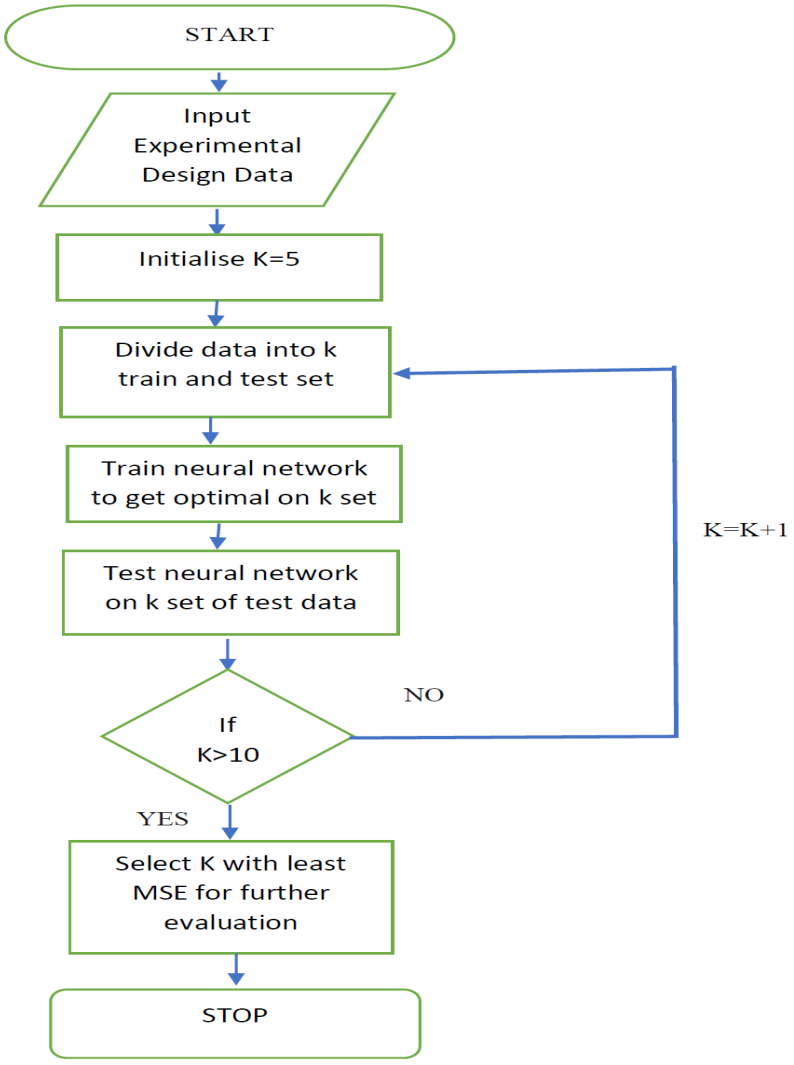
2.2 Artificial Neural Network

Multilayer perceptron ANN with one input layer, maximum of one hidden layer, and one output layer is used. Furthermore, to achieve the optimum topology of the network, different numbers of neurons in hidden layers are examined using a trial and error approach; while varying the number of neurons in the range 1-10. The learning is performed using Levenberg-Marquardt damped least-squares optimizer [19], this optimizer works on minimizing the least sum of squares function. By interpolating between Gauss-Newton [31] and Gradient Descent approach [32], the (non-negative) damping factor λ in the equation given by Levenberg-Marquardt (Eq.1) is adjusted at each iteration. If the reduction in “least sum of square function” is rapid, a smaller value of λ can be used, bringing optimizer closer to Gauss-Newton method. If an iteration gives

an insufficient reduction in residual (i.e., deviation from experimental data), λ can be augmented, giving a stride closer to the gradient-descent route. The transfer function used for hidden layers and the output layer are assumed to be log-sigmoid and linear, the performance index is chosen through Eq.2. Figure 1(a) represents the general ANN model diagram. The detailed procedure for training and validating the neural network can be seen in the flowchart in Figure 1(b). The software used for Simulation is MATLAB (R2017b) a multi-paradigm numerical computing environment developed by MathWorks Inc.



(a) A general ANN model diagram



(b) Flowchart for ANN with K-fold cross-validation

Figure 1. Traditional ANN model and Flowchart for training ANN model along with K-fold cross-validation

Levenberg-Marquardt:

$$[J^T J + \lambda \text{diag}(J^T J)]\delta = J^T [y - f(\beta)] \quad \textbf{(1)}$$

Mean Squared Error (MSE):

$$\sum_{i=1}^{n} \frac{(y_P - y_0)^2}{n} \quad \textbf{(2)}$$

where y_P = predicted value

y_0 = observed value

n = no of observations

J = Jacobian Matrix whose i-th row is equal to J_i, where f and y are vectors with i-th component $f(x_i, \beta)$ and y_i respectively, the equation is a set of linear equations, which is solved for δ(epoch).

In point of fact, transfer functions are a significant part of an Artificial Neural Network to introduce non-linear behavior to the network. Various transfer functions are Sigmoid, Tanh and Relu (Rectified Linear Units), the advantages and disadvantages are listed in Table 1. List of training parameters in the study is illustrated in Table 2.

Table 1. List of advantages and disadvantages of activation functions

Activation Function	Advantage	Disadvantage
Sigmoid	Easy to understand and apply, Easy to train on small dataset [26]	Vanishing Gradient Problem, the output is not zero centered.
Tanh	The output is zero centered	Vanishing Gradient Problem, hard to train on small datasets.
Relu	Avoids and rectifies vanishing gradient problems.	Can only be used with a hidden layer, hard to train on small datasets need much data for learning non-linear behavior.

Table 2. List of training parameters in the study

Parameter	Value
Train Function	'trainlm'
Max Epochs	50
Performance goal	0.000000000000001
Transfer Function	'log-sigmoid'
Learning rate	0.1

In this study, sigmoid activation function will be used for hidden layer as it is easy to get the train on a small dataset and experimental design data involves less data, as shown by Khatti et al. [30]. Also, this study involves solving regression problem where outputs are in continuous form; a linear function is used in the output layer for this purpose as shown by Khatti et al. [30].

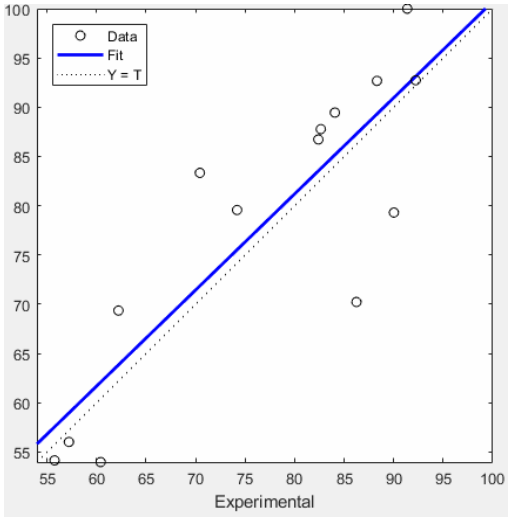
3 Results and Discussion

3.1 Case A - Modeling Remediation of Copper from Water Using Carboxylated Cellulose Nanowhiskers [29]

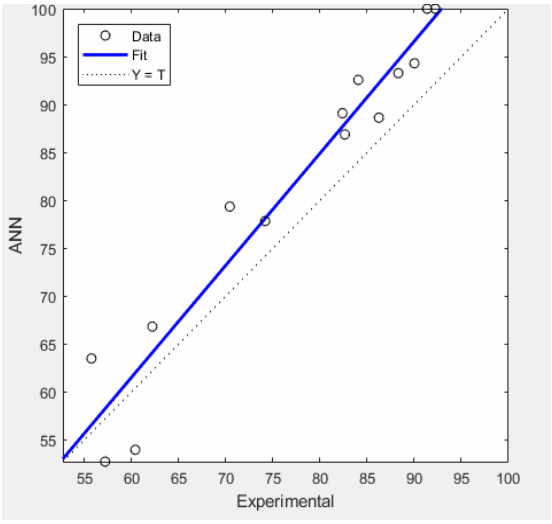
In this study, the capability of carboxylated cellulose nanowhiskers for remediation of copper from water is evaluated [29]. Input parameters include temperature 6 to 25℃, Initial Cu (II) ion concentration 10 to 60 mg/L and sorbent dosage 0.2 to 10 g/L. The output parameter is Cu (II) removal percentage. All values are coded in the dimensionless range of -2 to 2 as experimental design set. An ANN-based model with K-fold cross validation is developed for describing the removal of Cu (II) by modified CNWs [29]. Model is trained on a total of 20 experimental points, with Levenberg-Marquardt damped least square optimizer and K-fold cross-validation for validation purpose. To reduce deviations of predictions from experimental results, K was designated from 6-10. Since the values of K<, 6 generally offer high variance and value of K above 10 offers less bias on datasets, having less experimental design data. For each K, a number of neurons in the hidden layer is varied from 1 to 10. The correct value of K chosen for validation relies upon variance estimate of the dataset. High K reduces the variance estimate of the dataset. In contrast, in a lower value of K sufficient data is not included in the training set by the learner. Thus, the value of K chosen entirely depends on the design data. Average validation error for each configuration is noted (as shown in Table 3) and for this case, K=8 gives the best variance estimate. In which 8 different combinations of 16 train and 4 test sets, provide an intense check on 8 different test sets of the network with an optimal configuration of (4:8:1) MLP architecture. RSM and ANN model developed by Hamid et al. [29] shows good agreement on experimental data points. In order to have a comparison, accurate evaluation of models performance for 14 unseen experimental design data points of Cu (II) removal is evaluated by all three models in terms of R^2 (Figure 2). Corresponding residuals of each experimental data point by each model is calculated for comparison as shown in Table 4. From the results on unseen experiments, apparently residuals obtained from ANN chosen by K-fold cross-validation is the least. For example, on the first experimental point, it is 1.53 from the actual value which is better than other models (i.e., 7.14 from RSM and 4.81 from ANN by Hamid et al. [29]). In addition, R^2 for RSM model is calculated to be 0.73, whereas for ANN by Hamid et al. [29] evaluates to be 0.93 and for ANN after K-fold cross-validation on K=8 as 0.95 (Figure 2). These results confirmed that ANN models selected by K-fold cross-validation approach, unseen experimental design data are better than those made by RSM and ANN by

Table 3. Simulated results on experimental design data by Hamid et al. [29], where N1=input neuron, NH=Hidden Neuron, NO=Output neuron, K=Different combination of train and test sets, MSE=Average Mean-Square Error on the validation set

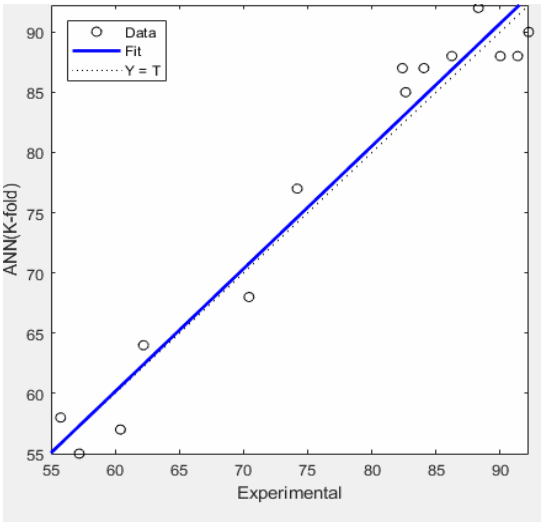
Runs	K	N1-NH-NO	MSE
1	6	4-2-1	10.5
2	6	4-4-1	7.2
3	6	4-6-1	4.3
4	6	4-8-1	3.6
5	6	4-10-1	3.6
6	8	4-2-1	7.3
7	8	4-4-1	6.2
8	8	4-6-1	2.2
9	8	4-8-1	0.9
10	8	4-10-1	0.9
11	10	4-2-1	9.3
12	10	4-4-1	4.8
13	10	4-6-1	3.9
14	10	4-8-1	2.1
15	10	4-10-1	1.9



(a) R^2=0.73



(b) R^2=0.9



(c) R^2=0.95

Figure 2. Comparison of model validation from (a) Prediction from RSM, (b) Prediction from ANN (developed by Hamid et al. [29]), and (c) Prediction from ANN (K-fold)

Table 4. Comparison of simulated results for 14 unseen Experiments, where T=Temperature in °C, C=Initial Cu (II) ion concentration, m=Sorbent dosage (g/L), Residual=Different between the predicted and actual value

No. of runs	T	C	m	Cu (II) removal (%)	% removed (RSM)	Residual	% removed ANN (Hamid et al. [29])	Residual	% removed ANN (K-fold validated)	Residual
1	21.3	50.3	5.1	62.21	69.35	-7.14	66.82	-4.81	63.74	-1.53
2	9.7	19.6	5.1	84.06	89.45	-5.39	92.57	-8.51	87.22	-3.16
3	15.5	35.0	8.1	82.39	86.72	-4.33	89.09	-6.7	86.88	-4.49
4	18	55.0	8.1	74.19	79.50	-5.37	77.83	-3.64	77.32	-3.13
5	10	55.0	4.0	60.41	53.97	6.44	53.96	6.45	57.54	2.87
6	20	35.0	9.5	82.64	87.77	-5.13	86.87	-4.23	85.62	-2.98
7	20	15.0	5.1	88.31	92.67	-4.36	93.27	-4.96	92.37	-4.06
8	6	10.0	8.1	91.37	99.98	-8.61	99.99	-8.62	88.92	2.45
9	25	60.0	8.1	70.43	83.33	-12.9	79.34	-8.91	68.24	2.19
10	25	60.0	4.0	57.20	56.01	1.19	52.72	4.48	54.73	2.47
11	10	10.0	4.0	92.23	92.73	-0.50	99.99	-7.77	89.88	2.35
12	6	35.0	2.1	55.75	54.14	1.61	63.49	-7.74	58.27	-2.52
13	6	20.0	2.1	86.24	70.22	16.02	88.62	-2.38	88.71	-2.47
14	10	10.0	2.1	90.02	79.30	10.72	94.30	-4.28	87.88	2.14

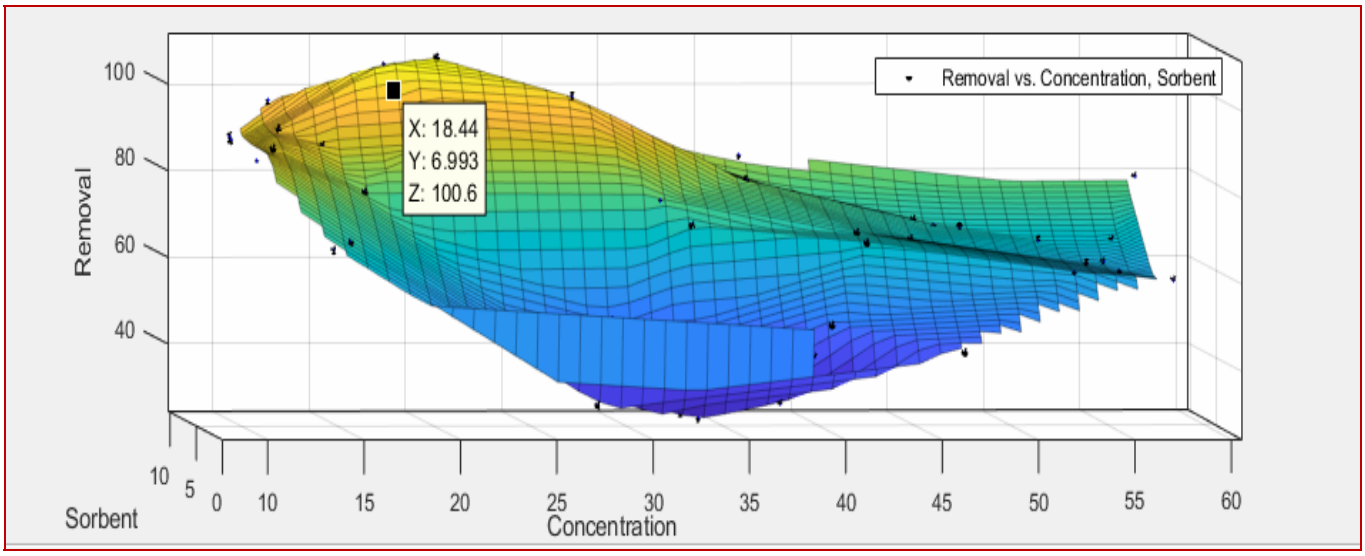


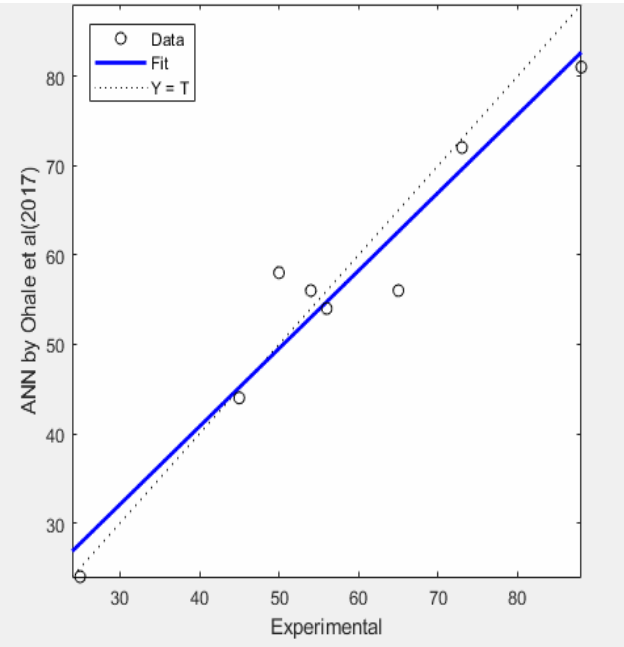
Figure 3. Response surface of interactive behavior of Concentration and Sorbent dosage on Cu (II) removal

Hamid et al. [29]. This is due to the fact of intense check on test cases performed in k-fold cross validation before selecting an appropriate model. By varying initial concentration from 10 to 60 (mg/L) and sorbent dosage from 0.2 to10 at fixed temperature 15.5°C, predicted results by ANN is shown in Figure 3. The percentage removal of Cu (II) increased on increase of sorbent dosage up to 7 g/L and then remains constant with an increase of sorbent dosage. This shows the high removal of Cu (II) at the high amount of sorbent dosage and low initial Cu (II) ion concentration. The results obtained by ANN are similar with findings reported in the literature [36]. A prediction of 100% removal matches well with that shown by Hamid et al. [29]. Hence, K-fold cross-validated ANN using Levenberg-Marquardt optimizer can be efficiently used in process modeling for experimental design [37].

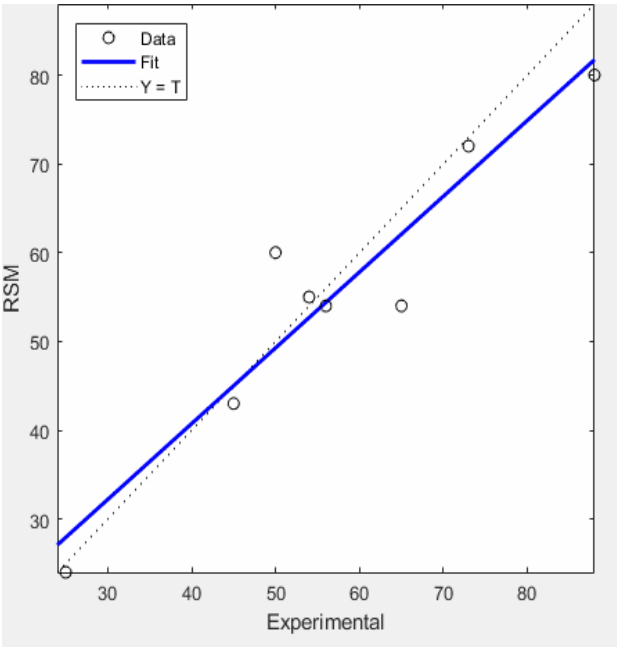
3.2 Case B-Modeling Alumina Dissolution from Azaraegbelu Clay [27]

As Ohale et al. [27] studied, input variables of experimental data included temperature 55-100 ℃, stirring speed 150-900 rpm, clay to acid dosage 0.01-0.1 g/ml, leaching time 40-250 min and leachant concentration 0.5-5.5 M. Alumina yield in % was the response variable. All input parameters are dimensionless encoded between the level of -2 to 2 according to experimental design. The process aims for optimal factor evaluation upon the dissolution of alumina from Azaraegbelu clay in acid solution. ANN with Levenberg-Marquardt optimizer and K-fold cross-validation technique is used to validate for the optimal model. Furthermore, to reduce deviations of prediction from experimental values, simulations are performed from K= 6 to 10. On every value of K (6-10) neurons are varied in the range from 1 to 10 to determine which configuration gives minimum deviations from

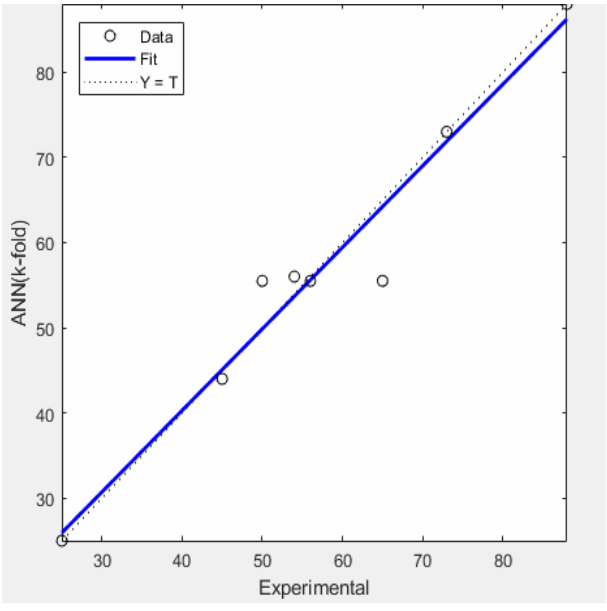
experimental results on validation data according to Kohavi [33]. To select the optimal configuration of ANN, number of neurons in the hidden layer is varied from 1 to 10 on each value of K. The best value of K obtained is K=10, in which network is trained and tested extensively on 10 different combinations of 29 train and 3 test sets, with obtained (5:8:1) optimal MLP architecture. The comparison of the results obtained by K-fold cross-validated ANN, RSM and ANN developed by Ohale at al [27] on a validation set of 8 experiments is shown in Figure 4. Regression plots of all three models show better correspondence with experimental data.



(a) R^2=0.89



(b) R^2=0.92



(c) R^2=0.94

Figure 4. Comparison of model validation from (a) Prediction from RSM (b) Prediction from ANN by Ohale et al. [27] (c) Prediction from ANN (K-fold)

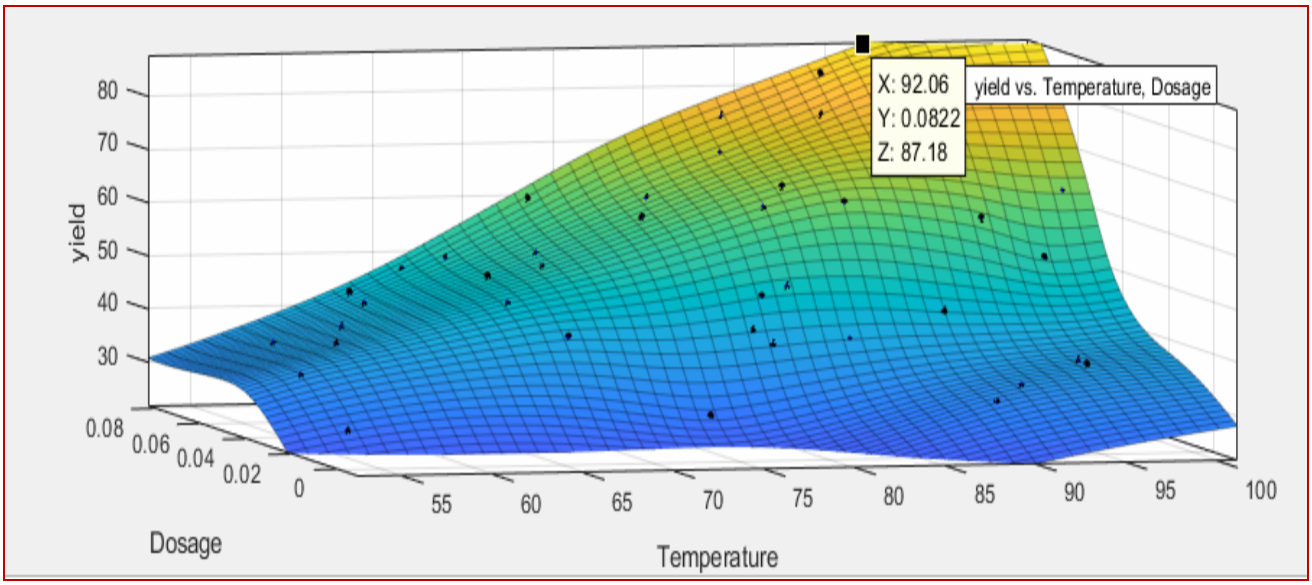


Figure 5. Response surface for the interactive behavior of temperature and dosage on alumina yield

Table 5. Simulated results for experimental design data by Ohale et al. [27], where N1=input neuron, NH=Hidden Neuron, NO=Output neurons, K= combinations of the train and test sets, MSE=Average Mean-Square Error on the validation set

Runs	K	N1-NH-NO	MSE
1	6	5-2-1	6.6
2	6	5-4-1	4.7
3	6	5-6-1	4.1
4	6	5-8-1	3.9
5	6	5-10-1	3.9
6	8	5-2-1	5.4
7	8	5-4-1	3.6
8	8	5-6-1	3.2
9	8	5-8-1	2.2
10	8	5-10-1	1.7
11	10	5-2-1	2.4
12	10	5-4-1	2.2
13	10	5-6-1	1.8
14	10	5-8-1	0.6
15	10	5-10-1	0.6

It can be seen from the plots that ANN obtained after K-fold (Figure 4(c)) cross-validation is capable of generalizing data more suitable with R^2=0.94 than ANN approach established by Ohale et al. [27] with R^2=0.92. Moreover, this is because ten different train and test sets produced in K-fold validation process help to check the model in a better way. However, Ohale et al. [27] only considered 1 set of train and test combination. In fact, the deviation can differ by changing train and test. Hence 1 set is of course not sufficient to generalize data in the better way. Thus, the optimal point in design space for getting maximum alumina yield % by analyzing interactive parameters is also studied, as shown in Figure 5. As literature [35] also pointed out, the consistency that increases in temperature substantially increases the dissolution rates of alumina solutes is indicated. A similar variation is observed in the case of ANN with K-fold cross-validation. By varying temperature and dosage and keeping other factors at the mean coded value of 0. Optimal solution predicted by developed ANN is at temperature=92.06 and dosage=0.08, for maximum yield of 87.18% (Figure 5). These results show excellent correspondence with the results as shown by Ohale at al [27] at temperature=92.5 and dosage=0.085 for maximum yield of 87%. Moreover, process response is maximized as temperature and dosage ratio is increased. A crucial observation in the analysis by ANN shows a continuous increase in alumina yield % as the temperature is increased. ANN can maximize alumina yield in the same and much better way as shown by RSM developed by Ohale et al. [27]. That is, this approach shows the better learning performed by ANN on significant temperature changes affecting the alumina yield. Table 5 shows the simulated results for experimental design data by Ohale et al. [27]

3.3 Case C-Modeling Electrospinning the Process of Polycaprolactone [30]

In this study, modeling of the electrospinning process of polycaprolactone is carried out for prediction of produced fiber diameter. The input parameters consists of polymer concentration from 8(%w/v) to 12(%w/v), voltage from 15(kV) to 21(kV) and Tip to collector distance from 8 cm to 12 cm. All dimensionless values are encoded from -1 to +1 according to experimental design data by Khatti et al. [30]. An ANN-based model while validating using K-fold cross validation is developed for predicting the fiber diameter produced through electrospinning process. Model is trained on a total of 17 experimental points as described by Khatti et al. [30] in the experimental design, with Levenberg-Marquardt damped least square optimizer and K-fold cross-validation. For validation purpose, K is ranged from 6-10 and for each K number of neurons in the hidden layer is varied from 1 to 10 to reduce deviations of predictions from experimental results. Average validation error for each configuration is noted in Table 6, and for this case, K=8 provides the most significant variance estimate. In which eight different combinations of 13 train and 4 test sets, make an appropriate check of ANN over eight different train and test sets with an optimal configuration of (3:8:1) MLP. Further, this helps to check ANN non-linearity over dataset in a generalized manner.

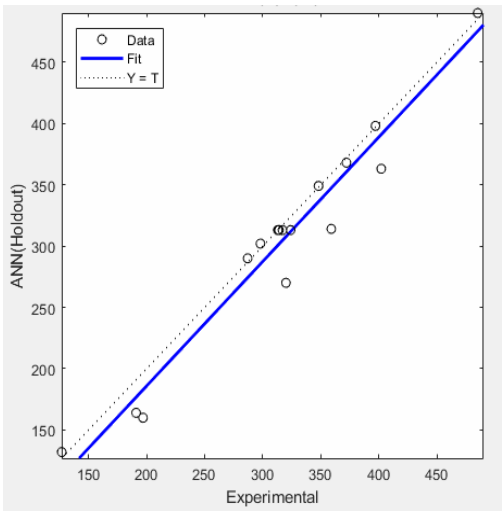
Table 6. Simulated results for experimental design data by Khatti et al. [30], where N1=input neuron, NH=Hidden Neuron, NO=Output neurons K= different combinations of the train and test sets, MSE=Average Mean-Square Error on the validation set

Runs	K	N1-NH-NO	MSE
1	6	3-2-1	6.2
2	6	3-4-1	5.8
3	6	3-6-1	5.8
4	6	3-8-1	4.3
5	6	3-10-1	3.2
6	8	3-2-1	4.4
7	8	3-4-1	3.6
8	8	3-6-1	2.2
9	8	3-8-1	1.1
10	8	3-10-1	1.1
11	10	3-2-1	5.1
12	10	3-4-1	3.9
13	10	3-6-1	3.8
14	10	3-8-1	1.8
15	10	3-10-1	1.3

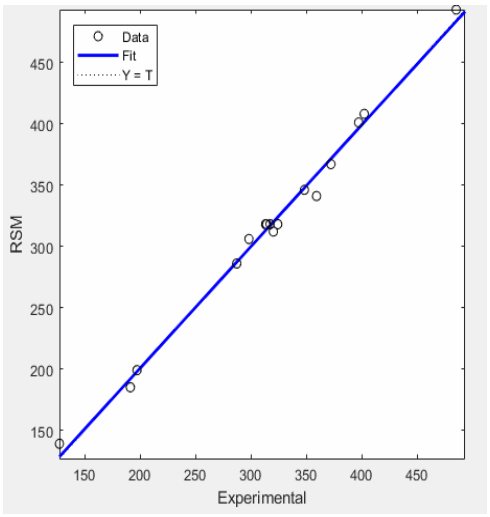
Apparently, RSM and ANN models developed by Khatti et al. [30] shows good correspondence with experimental data. Figure 6 shows a comparison of regression data of different obtained models. The R^2 of 0.94 by ANN validated by using K-fold cross-validation shows good correspondence on experimental design data. The advantage of this ANN model developed over that developed by Khatti et al. [30] is that this model is trained and tested in more combinations before simulations. As compared by ANN developed by Khatti et al. [30], training and testing different combination of dataset model can test model for non-linearity in data in a much better way. That is, results in better predictions by the model can be obtained. However, the disadvantage is that it involves much of computation as compared to the hold-out validation approach, as training and testing are completed on different combinations. In this case, ANN model developed by Khatti et al. [30] shows validation error of 0.88, and this can be improved by choosing a model with K-fold cross-validation, as shown in the regression results (Figure 6). The analysis of single-level effects produced by developed ANN on produced fiber diameter can be seen in Figure 7. As indicated in Figure 7, a linear behavior between concentration and produced fiber diameter is predicted by ANN. Also, this shows concentration as a useful and significant parameter in determining the diameter of the produced fiber. This linear behavior predicted by ANN matches well with experimental behavior shown by Khatti et al. [30]. That is, ANN can capture the behavior of experimental design data well and can be used for further modeling of this process.

4 Conclusion

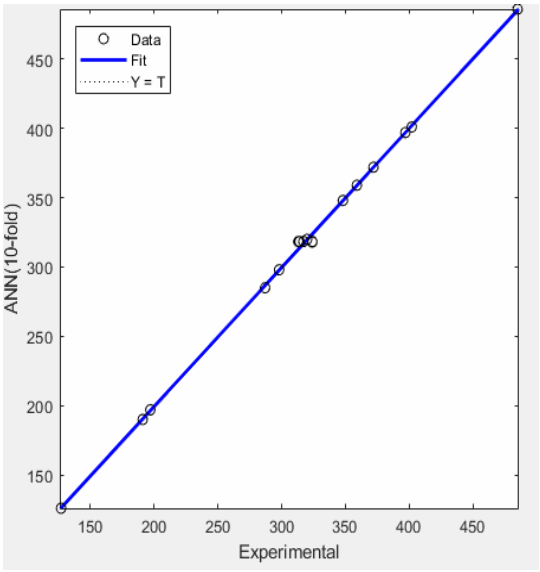
This work presents case-study on using Levenberg-Marquardt optimizer with K-fold cross-validation for training and validating ANN. These three sets of experimental design data are adopted from literature for evaluation and predictions. It is found that ANN with Levenberg-Marquardt optimizer while validating using K-fold cross-validation shows good correspondence with experimental design data. Moreover, increasing number of hidden neurons in the hidden layer overall error decreases in parallel. K-fold cross-validation allows using the entire dataset to train and test the obtained model. Although other modeling methodologies like RSM and ANN with hold-out validation can provide quality predictions(R^2) for the parameters within the design space, the ANN with Levenberg-Marquardt using K-fold cross-validation showed clear superiority over both in data fitting and estimation capabilities on data validation. Moreover, K-fold can split data into several train and test combinations which makes a model to get trained and tested on each experimental point. Meanwhile, having an average on K different test sets, it also helps to select a better optimal model which has better capability to generalize unseen data. Also, hold-out validation evaluation depends on which part of the data goes into the training set and which part goes in the test set. Performance evaluation will depend on the division of data which cannot guarantee generalization



(a) R^2=0.90



(b) R^2=0.88



(c) R^2=0.94

Figure 6. Comparison of model validation from (a) Prediction from RSM (b) Prediction from ANN (hold-out) (c) Prediction from ANN (K-fold)

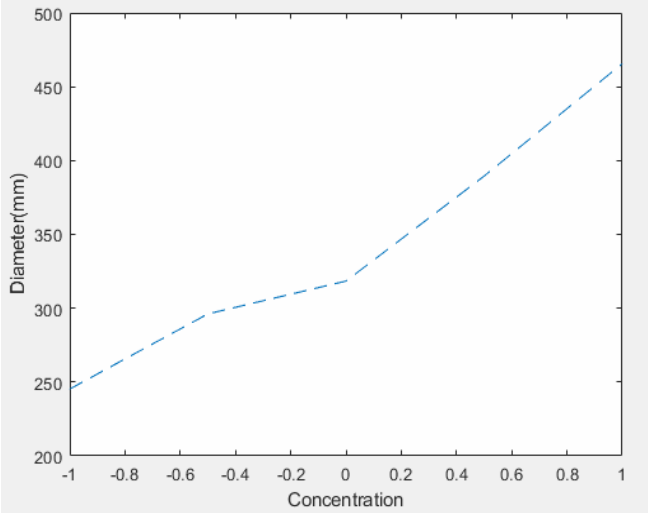


Figure 7. The effect of concentration on fiber diameter

on unseen data. Therefore, in this study the Levenberg-Marquart optimizer shows the superior capability of getting optimization over a small set of data. This approach is advantageous and novel for applications in experimental design data as they usually have limited data for training of neural network. In the future, the genetic programming [38-39] can be used for process optimization, since they can determine the optimized scenario for difficult multidimensional problems and it also makes use of nominal resources. The approaches based on genetic programming can determine several optimal solutions instead of a unitary solution.

Acknowledgements

The authors sincerely appreciate financial supports (MOST 106-2923-E-197-002-MY3) form the Ministry of Sciences and Technology, Taiwan for the study.

References

[1] G. E. P. Box, K. B. Wilson, On the Experimental Attainment of Optimum Conditions, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 13, No. 1, pp. 1-45, 1951.

[2] N.M. Anoop Krishnan, Sujith Mangalathu, Morten M. Smedskjaer, Adama Tandia, Henry Burton, Mathieu Bauchy, Predicting the Dissolution Kinetics of Silicate Glasses Using Machine Learning, *Journal of Non-Crystalline Solids*, Vol. 487, pp. 37-45, May, 2018.

[3] F. Lauer, Mlweb: A Toolkit or Machine Learning on The Web, *Neurocomputing*, Vol. 282, pp. 74-77, March, 2018.

[4] M. L. Giger, Machine Learning in Medical Imaging, *Journal of the American College of Radiology*, Vol. 15, No. 3, Part B, pp. 512-520, March, 2018.

[5] S. Bahaadini, V. Noroozi, N. Rohani, S. Coughlin, M. Zevin, J. R. Smith, V. Kalogera, A. Katsaggelos, Machine Learning for Gravity Spy: Glitch Classification and Dataset, *Information Sciences*, Vol. 444, pp. 172-186, May, 2018.

[6] I. Podolak, A. Roman, M. Szykuła, B. Zieliński, A Machine Learning Approach to Synchronization of Automata, *Expert Systems with Applications*, Vol. 97, pp. 357-371, May, 2018.

[7] A. Topîrceanu, G. Grosseck, Decision Tree Learning Used for the Classification of Student Archetypes in Online Courses, *KES2017- 21st International Conference on Knowledge Based and Intelligent Information and Engineering Systems*, Marseilles, France, 2017, pp. 51-60.

[8] C. Song, Research of Association Rule Algorithm Based on Data Mining, *2016 IEEE International Conference on Big Data Analysis (ICBDA)*, Hangzhou, China, 2016, pp. 1-4.

[9] S. Chen, J. Mao, F. Chen, P. Hou, Yong Li, Development of ANN Model for Depth Prediction of Vertical Ground Heat Exchanger, *International Journal of Heat and Mass Transfer*, Vol. 117, pp. 617-626, February, 2018.

[10] J. Schmidhuber, Deep Learning in Neural Networks: An Overview, *Neural Networks*, Vol. 61, pp. 85-117, January, 2015.

[11] M. Serrurier, H. Prade, Improving Inductive Logic Programming by Using Simulated Annealing, *Information Sciences*, Vol. 78, No. 6, pp. 1423-1441, March, 2008.

[12] Q. Shubo, G. Shuai, Z. Tongxing, Research on Paper Defects Recognition Based on SVM, *2010 WASE International Conference on Information Engineering*, Beidaihe, Hebei, 2010, pp. 177-180.

[13] M. Hong, K. Jing, L. Li-xiong, Research on Clustering Algorithms of Data Streams, *2010 2nd IEEE International Conference on Information Management and Engineering*, Chengdu, China, 2010, pp. 1-4.

[14] Y. Wang, H. Gao, G. Chen, Predictive Complex Event Processing Based on Evolving Bayesian Networks, *Pattern Recognition Letters*, Vol. 105, pp. 207-216, April, 2018.

[15] Y. Zhan, H. B. Ammar, M. E. Taylor, Scalable Lifelong Reinforcement Learning, *Pattern Recognition*, Vol. 72, pp. 407-418, December, 2017.

[16] G. Zhong, L. Wang, X. Ling, J. Dong, An Overview on Data Representation Learning: From Traditional Feature Learning to Recent Deep Learning, *The Journal of Finance and Data Science*, Vol. 2, No. 4, pp. 265-278, December, 2016.

[17] H. Dong, P. Lu, S. Zhong, C. Liu, Yi Ji, S. Gong, Person Re-identification by Enhanced Local Maximal Occurrence Representation and Generalized Similarity Metric Learning, *Neurocomputing*, Vol. 307, pp. 25-37, April, 2018.

[18] T. Zhang, F. Porikli, Cascade Residuals Guided Nonlinear Dictionary Learning, *Computer Vision and Image Understanding*, pp. 2-29, April, 2018.

[19] D. W. Marquardt, An Algorithm for Least-Squares Estimation of Nonlinear Parameters, *Journal of the Society for Industrial and Applied Mathematics*, Vol. 11, No. 2, pp. 431-441, 1963.

[20] Y. Kurtgoz, M. Karagoz, E. Deniz, Biogas Engine Performance Estimation Using ANN, *Engineering Science and Technology, an International Journal*, Vol. 20, No. 6, pp. 1563-1570, December, 2017.

[21] G. P. Gupta, R. R. Nair, R Jeyanthi, An ANN based SpO2 Measurement for Clinical Management Systems, *1st International Conference on Power Engineering, Computing and Control*, Chennai, Indian, 2017, pp. 393-400.

[22] H. K. Ghritlahre, R. K. Prasad, Application of ANN Technique to Predict The Performance of Solar Collector Systems- A Review, *Renewable and Sustainable Energy Reviews*, Vol. 84, pp. 75-88, March, 2018.

[23] L. Bahmani, M. Aboonajmi, A. Arabhosseini, H. Mirsaeedghazi, ANN Modeling of Extraction Kinetics of Essential Oil From Tarragon Using Ultrasound Pre-Treatment, *Engineering in Agriculture, Environment and Food*, Vol. 11, No. 1, pp. 25-29, January, 2018.

[24] T. D. Dongale, K. P. Patil, S. R. Vanjare, A. R. Chavan, P. K. Gaikwad, R. K. Kamat, Modelling of Nanostructured Memristor Device Characteristics Using Artificial Neural Network (ANN), *Journal of Computational Science,* Vol. 11, pp. 82-90, November, 2015.

[25] D. Navamani Kartic, B.CH. Aditya Narayana, M. Arivazhagan, Removal of High Concentration of Sulfate from

Pigment Industry Effluent by Chemical Precipitation Using Barium Chloride: RSM and ANN Modeling Approach, *Journal of Environmental Management*, Vol. 206, pp. 69-76, January, 2018.

[26] M. R. Sabour, A. Amiri, Comparative Study of ANN and RSM for Simultaneous Optimization of Multiple Targets in Fenton Treatment of Landfill Leachate, *Waste Management*, Vol. 65, pp. 54-62, July, 2017.

[27] P. E. Ohale, C. F. Uzoh, O. D. Onukwuli, Optimal Factor Evaluation for the Dissolution of Alumina from Azaraegbelu Clay in Acid Solution Using RSM and ANN Comparative Analysis, *South African Journal of Chemical Engineering*, Vol. 24, pp. 43-54, December, 2017.

[28] T. Fushiki, Estimation of Prediction Error by Using K-Fold Cross-Validation, *Statistics and Computing*, Vol. 21, No. 2, pp. 137-146, April, 2011.

[29] H. A. Hamid, Y. Jenidi, W. Thielemans, C. Somerfield, R. L. Gomes, Predicting the Capability of Carboxylated Cellulose Nanowhiskers for The Remediation of Copper From Water Using Response Surface Methodology (RSM) and Artificial Neural Network (ANN) Models, *Industrial Crops and Products*, Vol. 93, pp. 108-120, December, 2016.

[30] T. Khatti, H. Naderi-Manesh, S. M. Kalantar, Application of ANN and RSM Techniques for Modeling Electrospinning Process of Polycaprolactone, *Neural Computing and Applications*, pp. 1-10, April, 2017.

[31] Á. A. Magreñán, I. K. Argyros, *A Contemporary Study of Iterative Methods*, Academic Press, 2018.

[32] X. Zhu, Z. Gao, An Efficient Gradient-based Model Selection Algorithm for Multi-output Least-squares Support Vector Regression Machines, *Pattern Recognition Letters*, Vol. 111, pp. 16-22, August, 2018.

[33] R. Kohavi, A Study Of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, *IJCAI'95 Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montreal, Quebec, 1995, pp. 1137-1143.

[34] R. O. Ajemba, O. D. Onukwuli, Determination Of The Optimum Dissolution Conditions Of Ukpor Clay In Hydrochloric Acid Using Response Surface Methodology, *International Journal of Engineering Research and Applications*, Vol. 2, No. 5, pp. 732-742, September-October, 2012.

[35] D. Bingöl, M. Hercan, S. Elevli, E. Kılıç, Comparison of the Results of Response Surface Methodology and Artificial Neural Network for The Biosorption of Lead Using Black Cumin, *Bioresource Technology*, Vol. 112, pp. 111-115, May, 2012.

[36] F. Geyikçi, E. Kılıç, S. Çoruh, S. Elevli, Modelling of Lead Adsorption from Industrial Sludge Leachate on Red Mud by Using RSM and ANN, *Chemical Engineering Journal*, Vol. 183, pp. 53-59, February, 2012.

[37] R. Ramesh, R. Jeya Shakila, B. Sivaraman, P. Ganesan, P. Velayutham, Optimization of the Gelatinization Conditions to Improve The Expansion and Crispiness of Fish Crackers Using RSM, *LWT*, Vol. 89, pp. 248-254, March, 2018.

[38] A. Garg, K. Shankhwar, D. Jiang, V. Vijayaraghavan, B. N. Panda, S. S. Panda, An Evolutionary Framework in Modelling of Multi-Output Characteristics of the Bone Drilling Process, *Neural Computing and Applications*, Vol. 29, No. 11, pp. 1233-1241, June, 2018.

[39] Y. Huang, L. Gao, Z. Yi, K. Tai, P. Kalita, P. Prapainainar, A. Garg, An Application of Evolutionary System Identification Algorithm in Modelling of Energy Production System, *Measurement*, Vol. 114, pp. 122 -131, January, 2018.

Biographies

Kathiravan Srinivasan, received his B.E. (ECE), M.E., (CS) and Ph.D., (ICE) from Anna University Chennai, India. He is presently working as an Associate Professor in the School of Information Technology and Engineering at Vellore Institute of Technology (VIT), Vellore, India. His research interests include Machine Learning, Image Processing, Communication Systems & Networks, Multimedia and Vehicular Networks.

Aswani Kumar Cherukuri is Professor of School of Information Technology and Engineering, Vellore Institute of Technology (VIT), Vellore, India. He holds a Ph.D. degree in Computer Science from VIT, India. He also possesses Bachelor's and Master's Degrees in Computer Science from Nagarjuna University, India. His research interests are Information Security and Machine learning.

P.M. Durai Raj Vincent received his B.E. and M.E. from Anna University, Chennai, India. He received his Ph.D., from VIT University Vellore. He is presently working as an Associate Professor in the School of Information Technology and Engineering at Vellore Institute of Technology (VIT), India. His current research interest includes Security, Machine Learning and Data Analytics.

Ashish Garg received his B.Tech Degree in Information Technology from Vellore Institute of Technology (VIT), Vellore, India. He is presently working in Wipro Digital, India as a project engineer, indulging mainly in tasks of machine learning and automating the existing systems. His interest areas includes Applications of machine learning and user interface designs.

Bor-Yann Chen is Distinguished Professor in Department of Chemical and Materials Engineering, National I-Lan University, Taiwan. He has published 150+ SCI-peer reviewed articles in reputed journals and has many National Awards (e.g., Professor Yen-Ping Shih Best Paper Awards of 2007, 2011, 2013 and 2016 from Taiwan Institute of Chemical Engineers).