Biocatalysis and Agricultural Biotechnology 18 (2019) 101046

Contents lists available at ScienceDirect

Biocatalysis and Agricultural Biotechnology

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

Comparison of response surface methodology and feedforward neural network modeling for polycaprolactone synthesis using enzymatic polymerization

Harshini Pakalapati[a], Senthil Kumar Arumugasamy[a,*], Mohammad Khalid[b]

[a] Department of Chemical and Environmental Engineering, Faculty of Engineering, University of Nottingham Malaysia Campus, Jalan Broga, 43500, Semenyih, Selangor, Malaysia

[b] Graphene & Advanced 2D Materials Group (GAMRG), School of Science & Technology, Sunway University, No. 5, Jalan Universiti, Bandar Sunway, 47500, SubangJaya, Selangor, Malaysia

ARTICLE INFO

Keywords:
Polycaprolactone (PCL)
Bio-polymerization
Ring opening polymerization
Response surface methodology
Artificial neural network modeling

ABSTRACT

This study highlights the optimisation of process parameters for the synthesis of a biodegradable polymer – polycaprolactone using response surface methodology and artificial neural networks. Temperature, time, mixing speed and solvent volume are the parameters considered for optimisation and polymer yield was chosen as response. The results obtained from RSM displays a good agreement with 3.4% percent deviation between predicted values and actual values. Further, feedforward neural network (FFNN) modeling is developed with five different training algorithms. Out of all, Levenberg-Marquardt training algorithm proved to be best with lowest MSE, MAE, MAPE values of 0.10, 0.18 and 0.02 respectively. Both the techniques have been successful in predicting the biopolymer yield. Coefficient of determination (R^2) and Absolute average deviation (AAD) value for RSM are obtained better proving RSM superior to ANN in this study.

1. Introduction

As in today's events, implementing environmentally friendly products is becoming more and more essential in order to maintain greener and safer world. In this context, biodegradable polymers are one such product, whose demand has increased phenomenally in recent years (Markit, 2018). Plastic regulations and bans led to the increased demand and market value of the biodegradable polymers. Besides, their potential applications in industrial, medical and agronomical fields adds to their high demand. Among many biodegradable polymers, polycaprolactone is one such polymer with wide range of applications namely tissue engineering, drug delivery etc (Labet and Thielemans, 2009; Woodruff and Hutmacher, 2010). Conventional synthesis of polycaprolactone involves ring-opening polymerization using chemicals and metals as catalyst. In recent advances, enzyme catalysis is more predominant in usage, to avoid toxic residues produced in conventional polymerization. Additionally, their activity in mild conditions, high specificity for substrates and natural origin makes them prudent (Gröger, 2018). Nevertheless, manufacturing biodegradable polymers using enzymes as catalyst in large scale still remains a problem.

Moreover, optimisation of process parameters in any production is important endeavor for the economy and feasibility of the process. Besides, they do play a prominent role in determining the product yield and quality. In order to improve the yield, specific conditions for parameters are need to be selected. However, this process will be taunting when it comes to industrial scale. Regular optimisation methods like single factor analysis is tiresome with less accuracy (Witek-Krowiak et al., 2014). Furthermore, enzyme catalyzed polymerization does not allow any wastage of materials. Therefore, it's important to choose a powerful technique to predict the optimised conditions with good accuracy. Given that, design of experiments (DOE), a new statistical approach is considered to be the efficient method to evaluate the relation between independent process variables and chosen responses along with determining the optimised conditions. In this study, the process parameters were optimised for synthesis of polycaprolactone via enzyme catalysis. The objective of this study is to predict the optimised conditions to achieve maximum yield by performing minimum number of experimental runs. Also, to understand the process i.e. the interaction of each parameters and their effect on polymer yield.

Artificial neural networks (ANN) is one of the most established modeling techniques in the recent past. ANN is an information

* Corresponding author.
E-mail address: SenthilKumar.Arumugasamy@nottingham.edu.my (S.K. Arumugasamy).

https://doi.org/10.1016/j.bcab.2019.101046
Received 3 December 2018; Received in revised form 8 February 2019; Accepted 12 February 2019
Available online 13 February 2019
1878-8181/

processing system that has certain performance characters in common with biological neural networks. Several researchers have used ANN for modeling various non-linear systems especially in the field of polymerization. To name a few, Verpoort et al., (2018), used artificial neural network to model and verify material properties. Youshia et al., (2017) developed artificial neural network model including polymer viscosity, contact angle and interfacial tension, as input with particle size and polydispersity index as output. Arumugasamy et al. (2012) used ANN for the ring-opening polymerization ε-caprolactone (ε-CL) using lipase Novozyme 435 as catalyst in flask level and reactor level. They used reaction time, temperature as the input parameters and molecular weight as the output for the flask studies whereas reactor impeller speed and molecular weight as the outputs for the reactor studies. Wong et al. (2018) used ANN for the bio-polymerization of ε-caprolactone, using lipase Novozyme 435 catalyst at varied reactor impeller speeds and reactor temperatures. Number average molecular weight (*Mn*), weight average molecular weight (*Mw*), and polydispersity index (PDI) estimation from bio-polymerization were included as the output. They developed a multilayer feedforward neural network (FFNN) model with 11 different training algorithms for the multivariable nonlinear bio-polymerization of polycaprolactone (PCL).

Several researchers around the world have compared RSM with ANN to identify the suitable modeling technique for their process. Yadav et al. (2018) used these two techniques for Oil agglomeration to recover fines and ultra-fines coal particles from the discarded tailing generated from coal preparation plants. They used, artificial neural network (ANN) and response surface methodology (RSM) modeling to predict the behaviour of coal oil agglomeration in terms of % ash rejection (% AR) and % combustible matter recovery (% CMR). They developed a three layered feed forward neural network by varying process variables namely solid concentration (SC), oil dosage (OD) and agglomeration time (AT). They found that ANN was better than RSM for their study. Aliakbarian et al. (2018) used RSM and ANN to determine the optimum conditions for spray drying microencapsulation of olive pomace extract, in order to stabilize its phenolic compounds using maltodextrin as carrier material. They proved that RSM was better than that of ANN for their study. Jha et al. (2017) used RSM and ANN in an Upflow Anaerobic Sludge Blanket (UASB) bioreactor for optimisation of hydrogen yield and COD (Chemical Oxygen Demand) removal efficiency. Their results suggested that ANN is a better model to predict the results with greater accuracy and higher reliability for the bioprocess parameter interactions associated to the fermentation process. Esonye et al. (2019) used RSM and ANN for biodiesel production from Swe*et al*mond (Prunus amygdalus) Seed oil (SASO) using through base (NaOH) transesterification. Their results of ANN and RSM models are observed to be comparable.

In this study RSM and FFNN have been used to predict the polymer yield of the bio-polymerization process. The bio-polymerization process was conducted using a flask setup in order to optimise the parameters including reaction temperature, reaction time, reaction mixing speed and solvent volume. The experimental trials were obtained from RSM (D-optimal model). Once the trials were conducted the reponses were used to develop the RSM model. These experimental analysis results were used to develop the FFNN model for the bio-polymerization of ε-caprolactone (ε-CL) to PCL. A total of 5 different training algorithms namely, Levenberg-Marquardt, Gradient descent backpropagation, BFGS quasi-Newton backpropagation, Conjugate gradient backpropagation with Fletcher-Reeves updates, Resilient backpropagation were evaluated. The dataset collected from 25 samples from the bio-polymerization experiments was used for the evaluation. Polymer yield estimation from bio-polymerization using different training algorithms are compared in term of estimation errors, which are mean absolute error (MAE), mean square error (MSE), and mean absolute percentage error (MAPE) with the experimental results. From these results the training algorithm that gave the better results was compared with RSM results to determine the most appropriate modeling technique for the bio-polymerization process.

2. Materials and methods

2.1. Materials

Caprolactone was used as monomer, Solvent opted for this bio-polymerization process was toluene. Lipase B from *Candida Antarctica* (Novozyme 435) has been used as a catalyst, to convert monomer to polymer. Other solvents like chloroform and methanol were used to terminate the process and precipitate the polymer respectively. Monomer and solvents have been obtained from Merck Sdn Bhd, Malaysia. Enzyme catalyst was purchased from Sigma Aldrich Sdn Bhd, Malaysia. All the materials were used as received. Solvents purchased are all of analytical grade.

2.2. Methods

2.2.1. Procedure for synthesis of polycaprolactone using Novozyme 435

Bio polymerization reaction was performed in a 250 ml conical flask with constant stirring using magnetic hot plate. Initially 10 ml of monomer and 1 g of enzyme are added to the flask. Throughout the process, the amount of monomer and enzyme were kept constant. Whereas, the temperature, reaction time, mixing speed (RPM) and solvent volume have been operated according to the DOE table obtained from the software. Later, when the reaction reached the specified time, termination of the reaction was done by adding excess amounts of cold chloroform. Furthermore, the enzyme filtration was done using filter paper, followed by rotatory evaporation in order to get rid of excess chloroform. Finally, methanol was added to precipitate the polymer and later dried in oven for overnight, maintained at 50 °C. Final dry polymer obtained was weighed and further monomer conversion percentage was calculated using gravimetric analysis.

2.2.2. D- optimal experimental design

Stat-ease software version 9 was used to develop the model. D-optimal design, one among the other designs in response surface methodology was opted to optimise the synthesis process to achieve maximum polymer yield. Four process parameters considered for optimisation are temperature (50–110 °C), reaction time (1–7 h), mixing speed (50–500 rpm) and solvent volume (10–60 ml). Polymer yield was chosen as response and range of parameters were chosen from previous literature (Kumar and Gross, 2000). Based on the 25 experimental runs generated by the optimal design, the effect of process parameters on response is studied and determined statistically. Four process parameters have been considered as continuous factors from low to high ranges.

The polynomial equation developed by the optimal design software enunciates the relation between process parameters and their chosen responses. Represented in equation (1)

$$Y = \beta_0 + \sum \beta i\, Xi + \sum \beta ij\, Xi\, Xj + \sum \beta i Xi^2 \quad (1)$$

Where Y denotes the response, β0 is coefficient of independent constant, βi is the individual coefficient, Xi stands for independent parameters (i = A, B, C, D), βij is for interaction coefficient effects (i = A, B, C, D and j = A, B, C, D).

Significance of the model was given by Analysis of variance and data was analysed through analysis of variance i.e. ANOVA, to investigate the importance of each process parameters along with their interactions. Statistical significance and the quality of the model fit were expressed by *F*-test and the p value along with few more terms explained in later sections.

2.3. Gravimetric analysis

The monomer conversion percentage was reported using gravimetric analysis. The conversion rate was calculated gravimetrically as shown in equation (2).

$$\text{Monomer conversion } (\%) = \frac{WP}{WM} \times 100 \quad (2)$$

Where, WP is the weight of the polymer and WM is the weight of the monomer.

3. Artificial neural networks

Artificial neural network is an empirical modeling technique which is used to predict the experimental results by using the experimental operating parameters. The operating parameters are the inputs and experimental results are the outputs of the model. ANN model consists of 3 important layers of neurons. The first layer being the “input layer”, the second layer is “hidden layer” and the third layer is the “output layer”. Each of these layer has information stored in “nodes” or “neurons”. The input layer has the information of the operating parameters stored in the “input neurons”. This information is transferred to the next layer ie. the hidden layer where the processing takes place. The processing step is as follows: The scalar input p is multiplied by the scalar weight w to form wp, one of the terms that is sent to the summer. The other input, 1, is multiplied by a bias b and then passed to the summer. The summer output n, often referred to as the net input, goes into a transfer function f, which produces the scalar neuron output a. Fig. 1 shows the single input neuron.

In this study, feedforward neural network modeling (FFNN) was used to predict the biopolymer (PCL) yield from the enzymatic polymerization process. FFNN is one of the most preferred types of ANN. The information transfer only in one direction, ie. the forward direction from input layer to the hidden layer and then from hidden layer to output layer.

The experimental results of bio polymer yield obtained from the thermogravimetric method was used as the output for the model and the operating conditions namely the reaction temperature, reaction time, mixing speed (RPM) and solvent volume were used as the inputs of the model. This FFNN model would be useful for predicting the biopolymer yield at any other operating conditions for which the experiments were not carried out provided the operating conditions are with in the mentioned ranges.

In this paper the MATLAB 2015a™ version was used to develop the codes. The steps in designing the FFNN model are mentioned below:

1 Data collection
2 Creation of network
3 Configuring the network
4 Initialling the weights and biases
5 Training the network
6 Testing the network

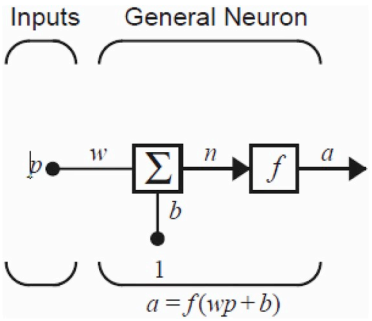

Fig. 1. Single input neurons.

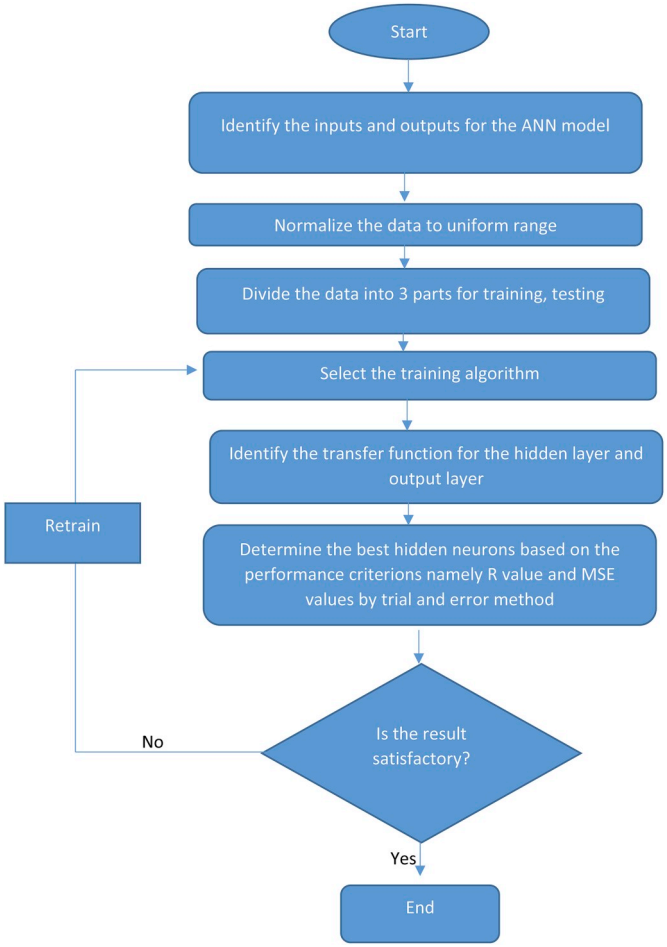

Fig. 2. Procedure to develop the ANN model with different training algorithms.

7 Validating the network

The stepwise procedure to develop the ANN model with different training algorithms is shown in Fig. 2.

3.1. Data collection and data division

The data is obtained from RSM trials obtained from the Design of Experiments (DOE) software for bio-polymerization process. Reaction temperature, reaction time, mixing speed and solvent volume which act as the inputs and bio-polymer yield acts as the outputs for the ANN model. The data is used to develop the FFNN model for the bio-polymerization process.

3.2. Data normalization

Normalizing the data is very important in ANN, as the inputs and output fall in different ranges. Normalizing the data makes sure the data falls in the same uniform range. In this study the data was normalized using the “zscore” function. This function normalizes the data with in the range of −1 and 1. This normalization helps in improving the prediction capability of the model.

3.3. Transfer function

The FFNN model consists of 1 hidden layer with tangent sigmoid transfer function (*tansig*) and the same at the output layer. Since there is only one output from the process ie. the bio-polymer yield and 4 inputs to the process, the model would be considered as a multi input single

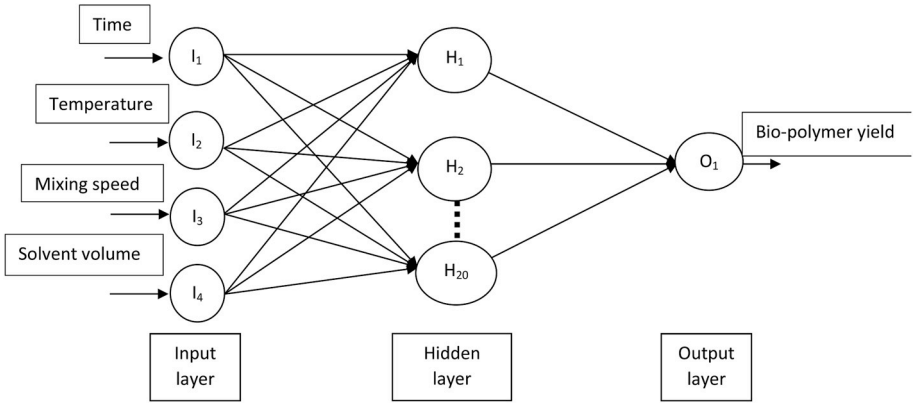

Fig. 3. Topology of a 3-layer ANN.

output model (MISO). The experimental output ie. the biopolymer yield will be considered as the target data for the FFNN modeling. The FFNN model structure used in this study is shown in Fig. 3.

Once the data is normalized, the inputs and outputs are divided into 3 parts namely training, testing and validation. The data is divided using the "dividend" function in MATLAB™. There are 25 experimental trials from the RSM. 15 data (60%) were chosen for training, 5 data for testing (20%) and 5 data for validation (20%).

3.3.1. Network training, validation and testing

Once the data is divided, the input and the output data chosen for "training" are used to train the neural network. This step is for the model to get trained with the available data (both input and output). This step is mandatory so that when a new set of data is introduced to the network without the experimental results, the model should be able to predict the results (output). Once the network is trained, the "testing" and "validation" of the model was carried out. This inturn shows whether the model has good prediction capability ie. a well training model is able to minimize the difference between the predicted and the experimental results which is know as "error".

The number of hidden neurons and number of hidden layers have a significant role to play inorder to minimize this "error". The hidden neurons are varied from 1 to 20. This is done by using the trial and error method and varying the hidden neurons in this range is to identify and test the hidden neuron that provides a good performance of the model ie. the model has the least "error".

For the network to get trained 5 training algorithms including Levenberg-Marquardt, Gradient descent backpropagation, BFGS quasi-Newton backpropagation, Conjugate gradient backpropagation with Fletcher-Reeves updates, Resilient backpropagation were preferred in this study. The "default" values of different training parameters namely the epochs, goals, learning rate etc for each of the training algorithms are the used in this study.

3.3.2. Optimisation of ANN structure

Once the network was trained with the training algorithms, the simulation was carried out so as to compare the predicted results with the experimental results. The network performance is determined using the mean squared error (MSE), mean absolute error (MAE) and mean abosulte percentage error (MAPE). The hidden neurons that provide the lowest MSE, MAE and MAPE values are considered to be the optimal values. The more these values are closer to zero better is the model. Another important performance criterion is the regression value (R). The closer the R value to 1, better is the model.

$$MSE = \frac{1}{n}\sum_{x=0}^{n}(Px - Ax)^2 \quad (3)$$

$$MAE = \frac{1}{n}\sum_{x=0}^{n}|Px - Ax| \quad (4)$$

$$MAE = \frac{1}{n}\sum_{x=0}^{n}\left(\frac{Ax - Px}{Ax}\right) \times 100 \quad (5)$$

$$AAD = \frac{100}{n}\sum_{i=1}^{n}\frac{Px - Ax}{Ax} \quad (6)$$

where n is the number of experiments; Px the experimental value for ith experiment; Ax is the predicted value for the ith experiment; Ye is the average of the experimental values.

4. Results and discussion

4.1. ANOVA

ANOVA evaluates the regression analysis of the model developed (Faraway, 2002). It consists of set of evaluation terms such as coefficient of determination (R^2), adjusted R^2, predicted R^2, coefficient of variance (C.V.%), adeq precision, F value and p value to explain the significance of the model. Table 1 presents all the values related to the developed model. Table 2 presents the experimental runs with actual and predicted values of the bio-polymer yield. The polynomial equation regenerated for the model was given in the below equation (3)

$$R1 = +5.86 + 0.4834*A - 0.0184*B - 0.0153*C + 0.8526 + D - 0.1779*AC - 0.2986*AD - 0.6345*BC - 0.5309 - 0.4687*CD + 1.17*A^2 + 0.4192*B^2 + 1.05*C^2 - 0.9685*D^2 \text{ (Eq 3)}$$

Whereas R1 represents the response that is monomer conversion percentage and A, B, C & D represents temperature, reaction time and solvent volume respectively.

From Table 1, the coefficient of determination (R^2) for response is 0.99 respectively. The R^2 values near to ≈ 1, indicates the model significance and also the real connection between the process parameters and responses (Gunawan et al., 2005). The values of predicted R^2 and adjusted R^2 are 0.94 and 0.97 respectively, they represent the high degree variations of the model which will help in predicting the

Table 1
Model analysis by ANOVA.

ANOVA terms	values
R2	0.9914
Adj R2	0.9774
Pre R2	0.9589
F value	70.82
P value	< 0.0001
CV	1.50
Adequate precision	26.9

Table 2
Experimental and predicted values by DOE and ANN.

S no	Temperature	Time	Mixing speed	Monomer/solevnt ratio	By D – optimal (DOE)		By ANN
					Experimental values	Predicted values	(train LM)
1	110	7	50	10	8.3	8.33	8.30
2	85	4	50	30	6.9	6.88	6.90
3	101	1	500	60	8.4	8.49	7.82
4	79	1	219	10	5.7	5.59	6.40
5	50	1	500	20	7.9	7.75	7.57
6	50	7	500	10	6.4	6.40	6.41
7	110	7	405	10	7.1	7.06	6.91
8	50	6	500	60	7	6.98	7.01
9	110	7	500	60	6.7	6.72	7.53
10	110	1	500	10	7.8	7.85	7.71
11	50	4	50	10	4.1	4.30	4.10
12	50	1	151	30	6.4	6.38	6.37
13	80	7	50	60	7.8	7.78	7.80
14	110	4	208	60	7.5	7.40	7.40
15	50	7	500	10	6.4	6.40	6.41
16	101	1	500	60	8.6	8.49	7.82
17	110	4	208	60	7.3	7.40	7.40
18	87	5	50	30	6.8	6.83	6.79
19	50	7	223	40	7.2	7.21	7.52
20	50	1	50	60	8.4	8.49	8.42
21	110	1	50	30	4.3	4.50	4.31
22	110	1	500	10	7.9	7.85	7.71
23	99	7	259	40	7.1	7.15	7.09
24	50	1	500	20	7.6	7.75	7.57
25	70	3	311	60	6.1	6.13	6.16

optimised conditions with good accuracy and also controlling the predictions to avoid any overfitting of the data. For the significant statistical model, the value of the difference between adjusted R^2 and predicted R^2 will be lesser than 0.2. From Table 1, the difference between these two terms is less than 0.02, presenting the significance of the model. It also explains the consistency of the variations considered throughout the model development. Low values of C.V.% indicates the great interrelationship and accuracy in-between the actual values and the predicted values. The CV value obtained from the developed model is 1.50 indicating high accuracy among the actual and predicted values of the polymer yield.

Other variables such as F and p values have been used to explain the model significance. F value is measure of scattered data from that of the mean data, indicating the two variance ratio. On the other hand, p value stands for the statistic indicating the possibility of observed F value are under acceptance of null hypothesis (Chaibakhsh et al., 2013). And the p values less than 0.05 indicates the denial of null hypothesis through the complete model (Sellke et al., 2001). F and p value for the response are 7.82 and < 0.0001 respectively, indicating the model terms are significant. Subsequently, observed huge F value could not occur due to noise. Besides, this is well supported by the non-significant value (0.94) obtained for the lack of fit.

Additionally apart from ANOVA results and their validation, software generates diagnostic plots to validate the model significance. Fig. 4 depicts the normal plot of the residual data, as all the points falls near to the line almost in straight line following the normal distribution. Fig. 5 shows residual experimental runs are assigned along the regression line representing good compliance between the predicted and actual values.

Furthermore, Figs. 6 and 7 illustrates the externally studentized residuals versus predicted and residuals vs predicted for the selected response respectively. Both of the plot are exhibiting the scattered residual runs within the brim of ± 4, supporting 95% of the confidence limits relying on the run number. Consequently, the irregular trend of the scattered runs, explains that the values predicted by the model, has both lower and higher values (than the actual runs) with the equal prospect (Pham, 2006).

Finally there is perturbation plot depicted in Fig. 8 explaining the

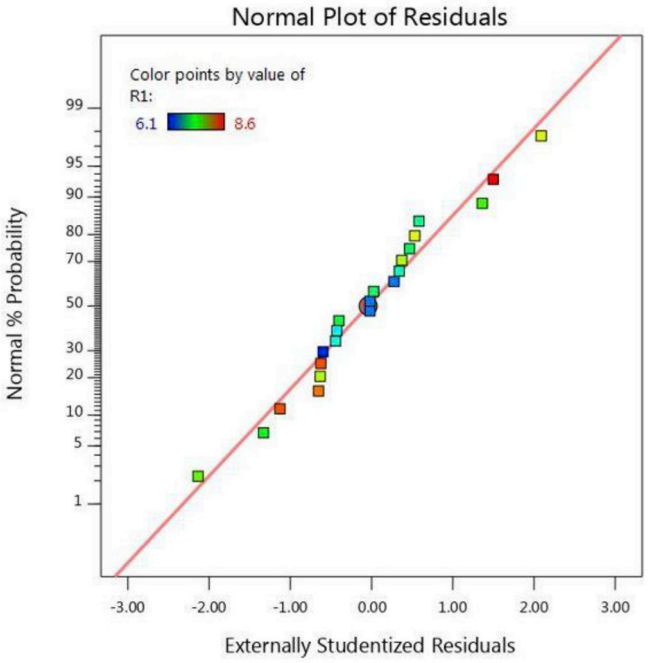

Fig. 4. Normal plot for the response (polymer yield).

effects of the four process parameters i.e. temperature, reaction time, mixing speed and solvent volume on the polymer yield. Slopes or steep curves in the perturbation curve speak for susceptibility of the chosen parameters on the polymer yield. Conversely, a flat line shows insensitivity to change in that particular factor. Thus, the perturbation plot could be used to find those factors that most affect the response (Sivakumar et al., 2007; Bashir et al., 2010). From Fig. 8, the slopes or steep deviations occurred depicts the sensitivity and profound effect of all the four process parameters on the polymer yield.

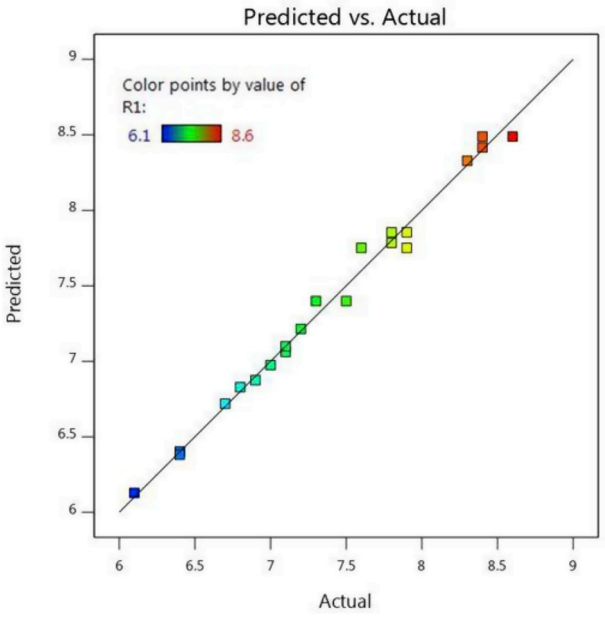

Fig. 5. Plot for Residual vs predicted runs.

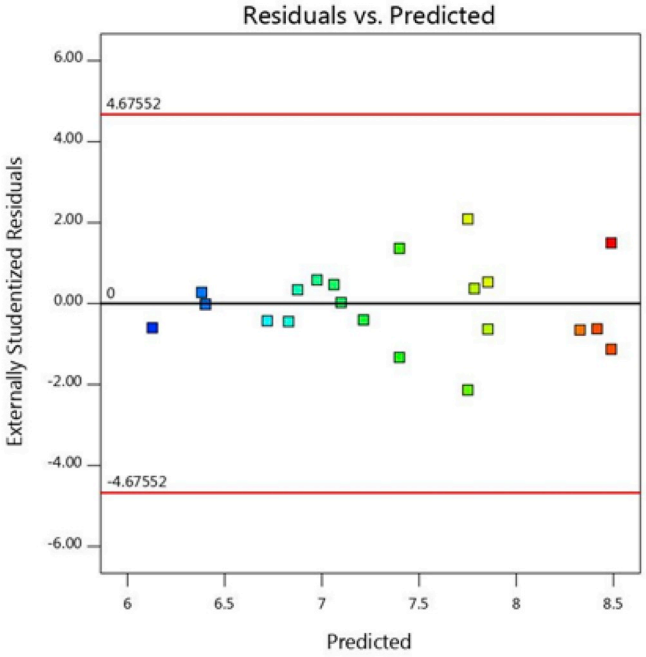

Fig. 7. Plot for Residuals vs Predicted.

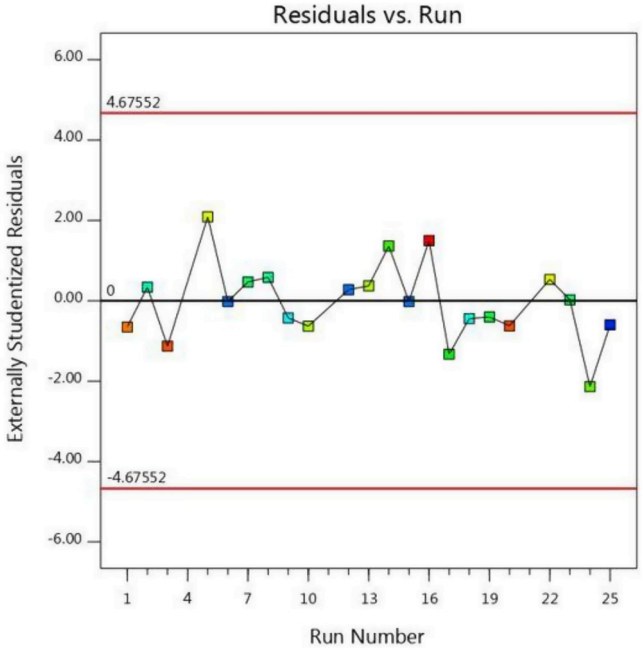

Fig. 6. Plot for Residual vs run.

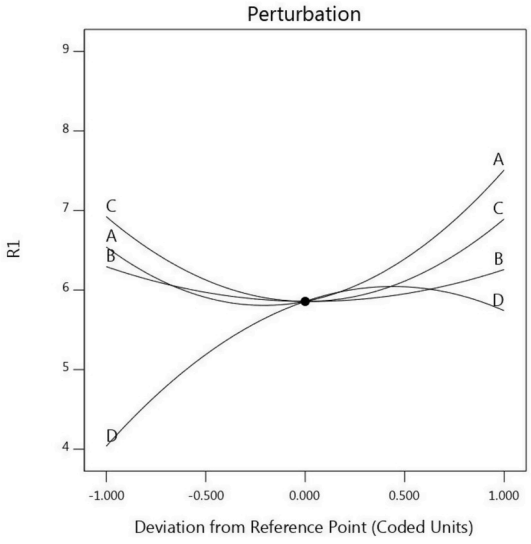

Fig. 8. Perturbation curve.

4.2. Validation of the developed model

Validation was performed by conducting the predicted optimised experimental run. Table 3 presents the predicted optimised values of process parameters and the observed experimental response with the deviation percentage. Up to date, most of the researchers reported the polycaprolactone synthesis using the single variate approach, where it takes 60–70 runs to optimise the process parameters. Thus, this study attempts to achieve the optimised conditions to achieve the maximum polymer yield by performing only 25 runs. Additionally, this developed model not only simplifies the process by reducing the number of trials but also reduces the usage of chemicals, which in turn reducing the production cost.

Therefore to perform the validation, the polycaprolactone was synthesised by employing the predicted conditions developed by the D-optimal design. From Table 3 it is evident that predicted and observed values of the experimental run are almost equivalent with the percentage error of 3.4%.

Table 3
Validation Of The Rsm Model.

	P1	P2	P3	P4	R1
Predicted	75.9	4	295	15	8.7
Observed	75	4	295	15	8.4
Error percentage					3.4%

*P1 = Temperature, P2 = Reaction time, P3 = mixing speed, P4 = Solvent volume, R1 = response (polymer yield).

4.2.1. Artificial neural networks

In this study 5 different training algorithms were used for ANN modeling which includes Levenberg-Marquardt, Gradient descent

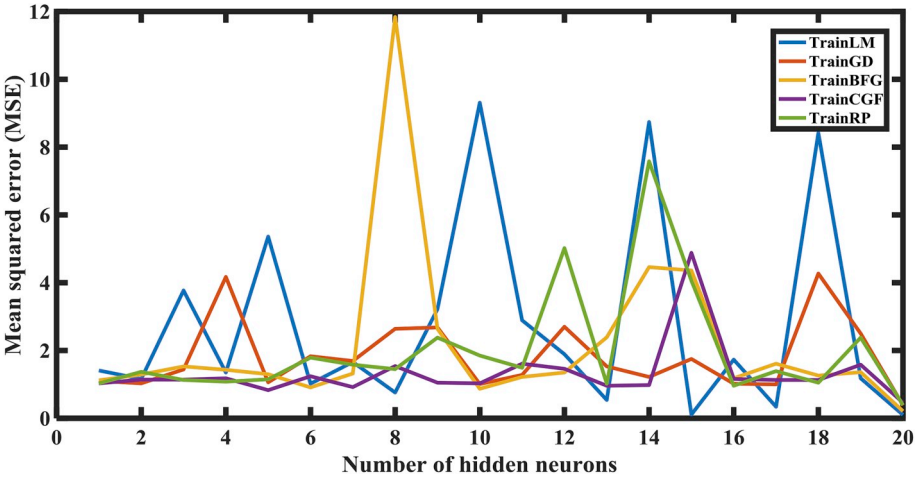

Fig. 9. Comparison of mean squared error (MSE) values produced via network training using different training algorithms.

backpropagation, BFGS quasi-Newton backpropagation, Conjugate gradient backpropagation with Fletcher-Reeves updates, Resilient backpropagation. The objective was to predict the PCL yield using these training algorithms. The most appropriate training algorithm was decided based on the performance criterions namely mean squared error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE) for the training, testing and validation data.

The data is obtained by carrying out the trails from RSM. The operating parameters namely the reaction time, reaction temperature and mixing speed, solvent volume act as the input and PCL yield acts as the output. The 25 data samples are divided into training, testing and validation. These data are used for developing the FFNN model with the five different training algorithms.

Figs. 9–11 shows the MSE, MAE and MAPE plots for the different hidden neurons (1 − 20) for the five different training algorithms chosen for this research. Fig. 12 shows the comparison between the experimental results and the ANN predicted results for the five training algorithms. The plot clearly shows that all the training algorithms have closely predicted the experimental results. Fig. 13 shows the comparison of R squared values produced via network training different training algorithms. Fig. 14 shows the error plots ie. the difference between the experimental results and the ANN predicted results for the five training algorithms. Fig. 15 shows the plot comparison between the experimental data and predicted data using Train Levenberg-Marquardt training algorithm which provides the most satisfying results.

Table 4 shows the MSE, MAE, MAPE values for each of the five training algorithm chosen in this study. For levenberg marqardt training algorithm, 20 hidden neurons were found to be the optimal value as the MSE value of 0.10 was the lowest. For Gradient descent backpropagation 20 hidden neurons was the optimal as the MSE value was 0.38. For BFGS quasi-Newton backpropagation lowest MSE value was 0.21 for 20 hidden neurons. For Conjugate gradient backpropagation with Fletcher-Reeves updates, 0.46 was the lowest MSE value for 20 hidden neurons. For Resilient backpropagation 0.39 was the lowest MSE. From Table 4 it is clear that all the five training algorithms show some poor performances ie. high values of MSE, MAE and MAPE for 1–9 hidden neurons. Among the five different training algorithms chosen for this study, “Trainlm” with 20 hidden neurons gives the lowest MSE, MAE and MAPE ie. 0.10, 0.18 and 0.02 respectively. Hence it shows that levenberg marquardt is the most appropriate training algorithms for this study.

From Table 4, it is seen that “trainlm” has the worst MSE value of 9.31 for 10 hidden neurons, “traingd” has a worst MSE value of 4.27 for 18 hidden neurons, “trainbfg” has the worst MSE of 11.84 for 8 hidden neurons, “traincgf” has worst MSE of 4.88 for 15 hidden neurons, “trainrp” has worst MSE of 7.58 for 14 hidden neurons. The worst performing training algorithms among all the five available is the “Trainbfg” which gives a MSE value of 11.84. “Trainrp” was the second worst giving a MSE value of 7.58. Even though “trainlm” gave a much higher MSE value of 9.31 at 10 hidden neurons, in the end it proved to be the best with the lowest MSE value.

Table 5 shows the MSE, MAE and MAPE values of the training data,

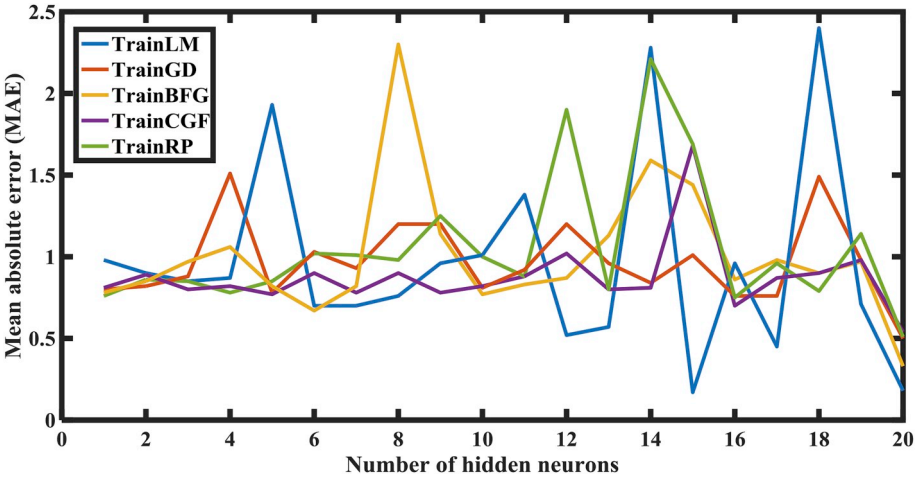

Fig. 10. Comparison of mean absolute error (MAE) values produced via network training using different training algorithms.

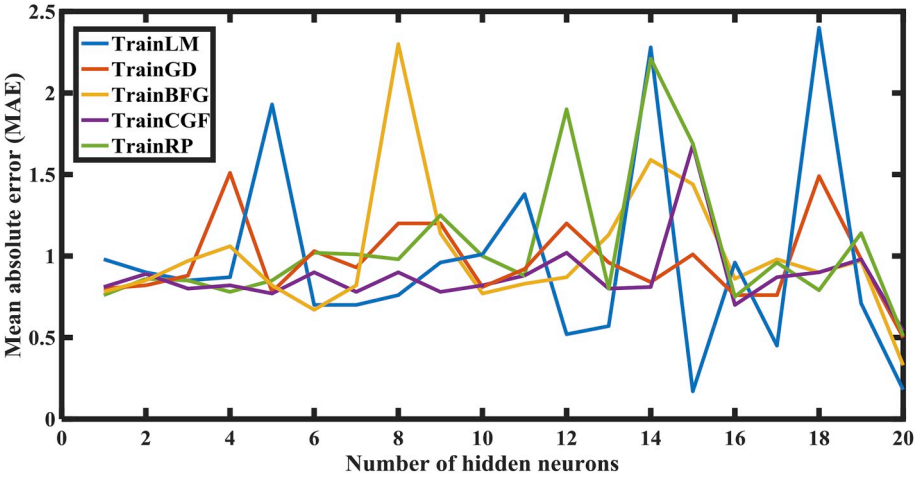

Fig. 11. Comparison of mean absolute percentage error (MAPE) values produced via network training using different training algorithms.

testing data and validation data of the neural networks individually. It can be seen that TrainLM show the least amount of MSE, MAE and MAPE among all the five training algorithms. Table 6 provides regression ® values for training, testing and validation for all the training algorithms.

4.2.2. Levenberg-Marquardt training algorithm

As shown in Table 4 and in Fig. 9, "trainlm" exhibits a large variability in its predictions capabilities with MSE values varying between 0.1 and 9.31. For "trainlm", 20 hidden neurons proved to be best value as it gave the lowest MSE, MAE, MAPE values of 0.10, 0.18 and 0.02 respectively. Hence it can seen that one of the worst performance of the network is also derived from "trainlm" giving MSE, MAE and MAPE values of 9.31, 1.01, 0.21 respectively at 10 hidden neurons. This is also shown in Figs. 9–11.

The values of best and worst prediction capability can be clearly seen in Fig. 12. It can be seen in Fig. 12 which compare the actual results with the predicted results that there is very little variation for the trainlm plot from the target plot as the error is quite minimal for the bio-polymer yield prediction. It can be seen clearly in Fig. 12 that the predicted results from trainlm has been successful in predicting the target values for most of the trails. Trainlm is often the fastest backpropagation algorithm in the toolbox, and is highly recommended as a first choice supervised algorithm, although it does require more memory than other algorithms.

4.2.3. Gradient descent backpropagation algorithm (traingd)

As can be seen in Figs. 9–11 and Table 4 the plot for MSE, MAE, MAPE using traingd fall in the range of 0.3–4.3, 0.5 to 1.5, 0.07 to 0.24 for the 1–20 hidden neurons. Out of this, the best values of 0.38, 0.5, 0.7 are obtained as MSE, MAE and MAPE respectively for the 20 hidden neurons. The worst values of MSE, MAE and MAPE are found to be 4.27 for hidden neuron 18, 1.51 for hidden neuron 4, and 0.24 for hidden neuron 18 respectively. The advantage of this method is that the gradient is accurate and the function converges fast. But when the training dataset is enormous, the evaluation of the gradient from all data points becomes expensive and the training time can be very long.

4.2.4. BFGS quasi-Newton backpropagation

From Figs. 9–11 and Table 4, it can be seen that the best values of MSE, MAE and MAPE are found to be 0.21, 0.33, 0.05 respectively at hidden neuron 20. The range for the MSE, MAE and MAPE are found to be 0.21–11.84, 0.33–2.3, and 0.05–0.34 respectively. The worst values of MSE, MAE and MAPE are found to be 11.84, 2.3, and 0.34 respectively for hidden neuron 8. The quasi-Newton method, trainbfg, is also quite fast similar to trainlm. Both of these methods tend to be less efficient for large networks (with thousands of weights), since they require more memory and more computation time for these cases.

4.2.5. Conjugate gradient backpropagation with Fletcher-Reeves updates (traincgf)

From Figs. 9–11 and Table 4, it can be seen the best values of MSE,

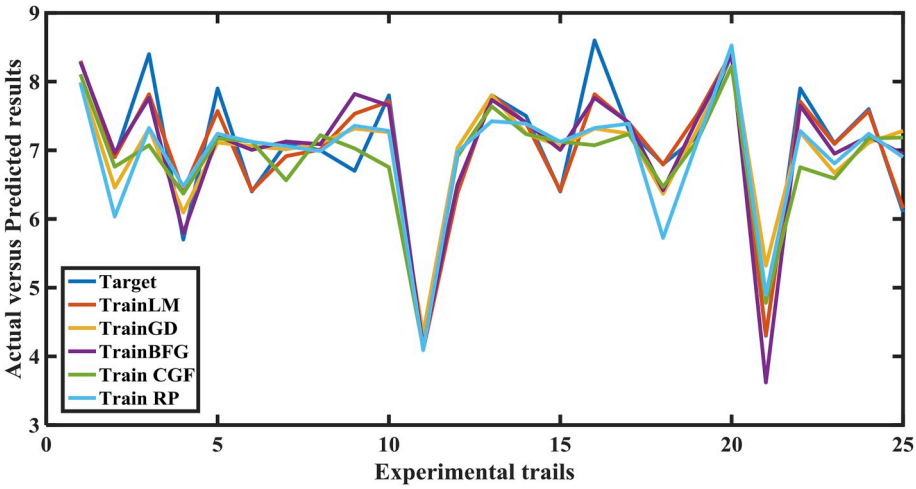

Fig. 12. Comparison of actual experimental results with the predicted values from different training algorithms.

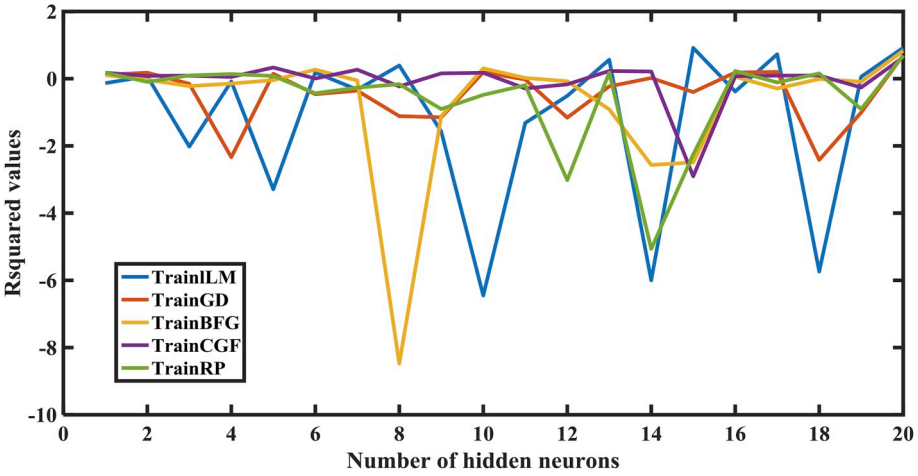

Fig. 13. Comparison of R squared values produced via network training different training algorithms.

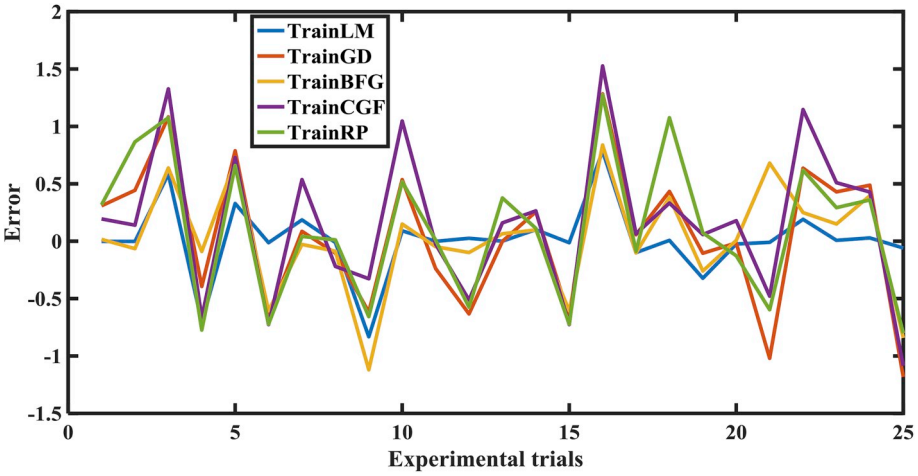

Fig. 14. Comparison of error values produced via network training different training algorithms.

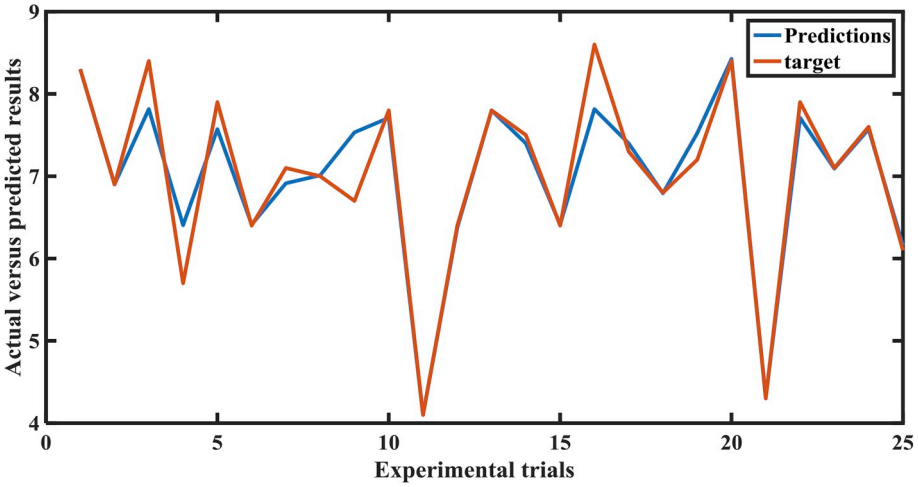

Fig. 15. Comparison of actual versus predicted results using Train Levenberg-Marquardt training algorithm (Best results).

MAE and MAPE are found to be 0.46, 0.54, 0.08 respectively at hidden neuron 20. The range for the MSE, MAE and MAPE are found to be 0.46–4.88, 0.54–1.68, and 0.08–0.26 respectively. The worst values of MSE, MAE and MAPE are found to be 4.88, 1.68, and 0.26 respectively for hidden neuron 15. Traincgf is a network training function that updates weight and bias values according to the conjugate gradient backpropagation with Fletcher-Reeves updates. The conjugate gradient algorithms are usually much faster than variable learning rate backpropagation, and are sometimes faster than trainrp (resilient backpropagation), although the results will vary from one problem to another. The conjugate gradient algorithms require only a little more storage than the simpler algorithms, so they are often a good choice for

Table 4
Table showing the MSE, MAE and MAPE values obtained from testing the ANN with Train Levenberg-Marquardt, Train Bayesian Regularization, Train Resilient Backpropagation and Train Scaled Conjugate Gradient.

Neurons	TrainLM			TrainGD			TrainBFG			TrainCGF			TrainRP		
	MSE	MAE	MAPE	MSE	MAE	MAPE	MSE	MAE	MAPE	MSE	MAE	MAPE	MSE	MAE	MAPE
1	1.41	0.98	0.14	1.10	0.80	0.13	1.12	0.78	0.13	1.03	0.81	0.13	1.03	0.76	0.12
2	1.16	0.90	0.14	1.03	0.82	0.13	1.29	0.85	0.14	1.14	0.89	0.14	1.37	0.86	0.15
3	3.77	0.85	0.11	1.44	0.88	0.15	1.53	0.97	0.15	1.14	0.80	0.13	1.13	0.85	0.13
4	1.35	0.87	0.15	4.17	1.51	0.20	1.43	1.06	0.16	1.18	0.82	0.13	1.08	0.78	0.13
5	5.36	1.93	0.29	1.06	0.78	0.13	1.30	0.82	0.13	0.83	0.77	0.12	1.15	0.85	0.14
6	1.02	0.70	0.12	1.83	1.03	0.15	0.91	0.67	0.11	1.24	0.90	0.14	1.79	1.02	0.16
7	1.66	0.70	0.10	1.69	0.93	0.14	1.32	0.82	0.14	0.92	0.78	0.13	1.58	1.01	0.15
8	0.76	0.76	0.11	2.64	1.20	0.19	11.84	2.30	0.34	1.54	0.90	0.15	1.45	0.98	0.16
9	3.22	0.96	0.13	2.68	1.20	0.17	2.64	1.14	0.19	1.05	0.78	0.13	2.38	1.25	0.21
10	9.31	1.01	0.21	1.01	0.81	0.13	0.87	0.77	0.12	1.03	0.82	0.13	1.85	1.00	0.15
11	2.89	1.38	0.20	1.28	0.92	0.14	1.22	0.83	0.13	1.61	0.88	0.12	1.49	0.88	0.15
12	1.89	0.52	0.07	2.70	1.20	0.19	1.35	0.87	0.13	1.46	1.02	0.15	5.02	1.90	0.27
13	0.54	0.57	0.08	1.53	0.96	0.15	2.39	1.13	0.18	0.96	0.80	0.12	1.02	0.80	0.13
14	8.74	2.28	0.33	1.22	0.84	0.14	4.46	1.59	0.22	0.98	0.81	0.12	7.58	2.21	0.31
15	0.11	0.17	0.02	1.75	1.01	0.15	4.36	1.44	0.21	4.88	1.68	0.26	4.06	1.69	0.24
16	1.73	0.96	0.15	1.02	0.76	0.12	1.18	0.86	0.14	1.16	0.70	0.11	0.96	0.75	0.12
17	0.34	0.45	0.07	1.00	0.76	0.12	1.61	0.98	0.15	1.13	0.87	0.13	1.39	0.96	0.14
18	8.42	2.40	0.35	4.27	1.49	0.24	1.26	0.90	0.14	1.13	0.90	0.13	1.05	0.79	0.12
19	1.18	0.71	0.12	2.50	0.98	0.14	1.36	0.97	0.14	1.58	0.98	0.16	2.38	1.14	0.18
20	0.10	0.18	0.02	0.38	0.50	0.07	0.21	0.33	0.05	0.46	0.54	0.08	0.39	0.51	0.07

Table 5
The MSE, MAE and MAPE for training, testing, and validation for Train Levenberg-Marquardt, Train Bayesian Regularization, Train Resilient Backpropagation and Train Scaled Conjugate Gradient training algorithms.

Training Yield		TrainLM	TrainGD	TrainBFG	TrainCGF	Train RP
	MSE	0.1129	0.2880	0.1967	0.3831	0.3515
	MAE	0.1930	0.4538	0.2940	0.5068	0.4960
	MAPE	0.0280	0.0656	0.0421	0.0713	0.0710
Testing yield		TrainLM	TrainGD	TrainBFG	TrainCGF	Train RP
	MSE	0.0082	0.6534	0.2830	0.6333	0.3192
	MAE	0.0591	0.7513	0.4637	0.7299	0.5337
	MAPE	0.0082	0.1273	0.0802	0.1125	0.0874
Validation yield		TrainLM	TrainGD	TrainBFG	TrainCGF	Train RP
	MSE	0.1460	0.3696	0.1853	0.4962	0.5626
	MAE	0.2477	0.3785	0.3179	0.4319	0.5275
	MAPE	0.0307	0.0474	0.0409	0.0529	0.0687

Table 6
R values for training, testing and validation for all the training algorithms.

	TrainLM	TrainGD	TrainBFG	TrainCGF	Train RP
Training	0.973	0.881	0.9215	0.923	0.884
Testing	0.916	0.634	0.970	0.322	0.350
validation	0.917	0.546	0.769	0.456	0.471

networks with a large number of weights.

4.2.6. Resilient backpropagation (trainrp)

From Figs. 9–11 and Table 4, it can be seen the best values of MSE, MAE and MAPE are found to be 0.39, 0.51, 0.07 respectively at hidden neuron 20. The range for the MSE, MAE and MAPE are found to be 0.39–7.58, 0.51–2.21, and 0.07–0.31 respectively. The worst values of MSE, MAE and MAPE are found to be 7.58, 2.21, and 0.31 respectively for hidden neuron 14. When training large networks, and when training pattern recognition networks, trainrp is good choices. Its memory requirements are relatively small, and yet they are much faster than standard gradient descent algorithms.

4.2.7. Comparison of ANN and RSM

In this study ANN and RSM modling techniques have been used to predict the PCL yield from the bio-polymerization process. Firstly RSM trails were obtained and the experiments were carried out. The analysed results were used to develop the RSM model. The experimental result were then used to develop th FFNN model using the different training algorithms. The hidden neurons were varied for each of these training algorithms based on the trial and error method. Levenberg marquardt training algorithm was found to be most appropriate one for this study with the hidden neuron of 20 giving the best result in terms of MSE and R values. The MSE value is 0.10 and R^2 value is 0.92. The experimental analysis results were used for developing the RSM model. The R^2 value is 0.99. The AAD values were used to compare the two model. The formula mentioned in equation (6) was used to calculate the AAD values. The AAD for RSM and ANN are 1.07 and 2.45 respectively. Comparing the R^2 values of RSM and ANN which are 0.99 and 0.92 respectively, it can be concluded that RSM is superior to ANN in predicting the PCL yield.

5. Conclusion

Enzymatic polymerization of ε-Caprolactone to polycaprolactone was carried out using *Candida Antartica Lipase* B in this study. Process parameters namely reaction time, reaction temperature, mixing speed and solvent volume were optimised. Successful implementation of D-

optimal for polycaprolactone synthesis was achieved. 75 °C, 4 h, 295 rpm and 15 ml of solvent volume were identified as optimised conditions to achieve the maximum polymer yield with 84% monomer conversion percentage. Also, all the four parameters chosen to have a profound effect on the response. Furthermore, error percentage is only 3.4% representing the robustness of the DOE. Prediction of PCL yield was carried out using the two modeling techniques RSM and FFNN. In this study, FFNN model was developed by comparing the 5 different training algorithms namely Levenberg-Marquardt, Gradient descent backpropagation, BFGS quasi-Newton backpropagation, Conjugate gradient backpropagation with Fletcher-Reeves updates, Resilient backpropagation. 25 experimental trials were used to develop the FFNN model and trainlm was found to be the most appropriate training algorithm with MSE, MAE and MAPE values being 0.10, 0.18, and 0.02 respectively. The comparison between RSM and ANN was shown using the AAD value. RSM having ADD value of 1.07 and ANN with 2.45 proved that RSM outperformed ANN in this study.

References

Aliakbarian, B., Sampaio, F.C., deFaria, J.T., Pitangui, C.G., Lovaglio, F., Casazza, A.A., Perego, P., 2018. Optimization of spray drying microencapsulation of olive pomace polyphenols using response surface methodology and artificial neural network. LWT - Food Sci. Technol. 93, 220–228. https://doi.org/10.1016/j.lwt.2018.03.048.

Arumugasamy, S.K., Uzir, M.H., Ahmad, Z., 2012. Modeling of polycaprolactone production from ε-caprolactone using neural network. In: Huang, T., Zeng, Z., Li, C., Leung, C.S. (Eds.), Neural information processing: 19th international conference, ICONIP 2012, Doha, Qatar, November 12–15, 2012, proceedings, part II. Springer, Berlin, Heidelberg, pp. 444–451. https://doi.org/10.1007/978-3-642-34481-7_54.

Bashir, M.J.K., Aziz, H.A., Yusoff, M.S., Adlan, M.N., 2010. Application of response surface methodology (RSM) for optimization of ammoniacal nitrogen removal from semi-aerobic landfill leachate using ion exchange resin. Desalination 254, 154–161. https://doi.org/10.1016/j.desal.2009.12.002.

Chaibakhsh, N., Abdul Rahman, M.B., Jesunathan, K., Basri, M., 2013. Optimization of biocatalytic synthesis of Chitosan Ester using response surface methodology. Chem. Solid Mater. 1, 41–55.

Esonye, C., Onukwuli, O.D., Ofoefule, A.U., 2019. Optimization of methyl ester production from Prunus Amygdalus seed oil using response surface methodology and Artificial Neural Networks. Renew. Energy 130, 61–72. https://doi.org/10.1016/j.renene.2018.06.036.

Faraway, J.J., 2002. Practical Regression and ANOVA Using R. University of Bath.

Gröger, H., 2018. Enzyme catalysis in the synthesis of pharmaceuticals. Bioorg. Med. Chem. 26, 1239–1240. https://doi.org/10.1016/j.bmc.2018.03.026.

Gunawan, E.R., Basri, M., Rahman, Mohd.B.Abd., Salleh, A.B., Rahman, R.N.Z.Abd., 2005. Study on response surface methodology (RSM) of lipase-catalyzed synthesis of palm-based wax ester. Enzym. Microb. Technol. 37, 739–744. https://doi.org/10.1016/j.enzmictec.2005.04.010.

Jha, P., Kana, E.B.G., Schmidt, S., 2017. Can artificial neural network and response surface methodology reliably predict hydrogen production and COD removal in an UASB bioreactor? J. Hydro. Energy 42, 18875–18883. https://doi.org/10.1016/j.ijhydene.2017.06.063.

Kumar, A., Gross, R.A., 2000. Candida antartica lipase B catalyzed polycaprolactone synthesis: effects of organic media and temperature. Biomacromolecules 1, 133–138. https://doi.org/10.1021/bm990510p.

Labet, M., Thielemans, W., 2009. Synthesis of polycaprolactone: a review. Chem. Soc. Rev. 38, 3484–3504. https://doi.org/10.1039/B820162P.

Markit, I., 2018. As Plastic Regulations and Bans Increase, Market Value for Biodegradable Polymers Exceeds $1 Billion and Will Rise Sharply by 2023. https://www.businesswire.com/news/home/20180726005162/en/Plastic-Regulations-Bans-Increase-Market-Biodegradable-Polymers, Accessed date: 26 July 2018.

Pham, H., 2006. Springer Handbook of Engineering Statistics. Springer Science & Business Mediahttps://doi.org/10.1007/978-1-84628-288-1.

Sellke, T., Bayarri, M., Berger, J.O., 2001. Calibration of ρ values for testing precise null hypotheses. Am. Statistician 55, 62–71. https://doi.org/10.1198/000313001300339950.

Sivakumar, T., Manavalan, R., Muralidharan, C., Valliappan, K., 2007. Multi-criteria decision making approach and experimental design as chemometric tools to optimize HPLC separation of domperidone and pantoprazole. J. Pharmaceut. Biomed. Anal. 43 (5), 1842–1848. https://doi.org/10.1016/j.jpba.2006.12.007.

Verpoort, P.C., MacDonald, P., Conduit, G.J., 2018. Materials data validation and imputation with an artificial neural network. Comput. Mater. Sci. 147, 176–185. https://doi.org/10.1016/j.commatsci.2018.02.002.

Witek-Krowiak, A., Chojnacka, K., Podstawczyk, D., Dawiec, A., Pokomeda, K., 2014. Application of response surface methodology and artificial neural network methods in modelling and optimization of biosorption process. Bioresour. Technol. 160, 150–160. https://doi.org/10.1016/j.biortech.2014.01.021.

Wong, Y.J., Arumugasamy, S.K., Jewaratnam, J., 2018. Performance comparison of feedforward neural network training algorithms in modeling for synthesis of polycaprolactone via biopolymerization. Clean Technol. Environ. Policy 20, 1971–1986. https://doi.org/10.1007/s10098-018-1577-4.

Woodruff, M.A., Hutmacher, D.W., 2010. The return of a forgotten polymer—polycaprolactone in the 21st century. Prog. Polym. Sci. 35, 1217–1256. https://doi.org/10.1016/j.progpolymsci.2010.04.002.

Yadav, A.M., Chaurasia, R.C., Suresh, N., Gajbhiye, P., 2018. Application of artificial neural networks and response surface methodology approaches for the prediction of oil agglomeration process. Fuel 220, 826–836. https://doi.org/10.1016/j.fuel.2018.02.040. 2018.

Youshia, J., Ali, M.E., Lamprecht, A., 2017. Artificial neural network based particle size prediction of polymeric nanoparticles. Eur. J. Pharm. Biopharm. 119, 333–342. https://doi.org/10.1016/j.ejpb.2017.06.030.