Indian Journal of Biotechnology
Vol. 18, July 2019, pp 269-279

Response surface and artificial neural network simulation for process design to produce L-lysine by *Corynebacterium glutamicum* NCIM 2168

Ashutosh Kumar Pandey[a*], Kritika Pandey[a], Ashok Pandey[c], Vivek Kumar Morya[b] and Lalit Kumar Singh[a]*

[a]Department of Biochemical Engineering, School of Chemical Technology, Harcourt Butler Technical University, Kanpur -208002, Uttar Pradesh (UP), India

[b]Centre for Energy and Environmental Sustainability, Lucknow-226 029, UP, India

[c]Centre for Innovation and Translational Research, CSIR-Indian Institute of Toxicology Research, Lucknow, India

Received 31 June 2019; revised 27 March 2019; accepted 3 May 2019

The L-lysine is one of the most important essential amino acid used in food and pharmaceutical industries. The present investigation was conducted to optimize the L-lysine production by *Corynebacterium glutamicum* (NCIM 2168). The production parameters such as the temperature, pH and glucose concentration (g/l) were optimised and evaluated by simulation method to develop a suitable model. The experimental design was done using central composite design (CCD). Total 20 set of experiments were performed according to the CCD. The factors and their responses were analysed by using the statistical tools: response surface methodology (RSM) and artificial neural network (ANN) linked with genetic algorithm (GA). The predicted optimum production of L-lysine was 19.003 g/l and 28.363 g/l by CCD-RSM and ANN-GA respectively. During validation by GA under optimized conditions, the L-lysine production was found to be 27.25 ± 1.15 g/l, which was significantly high than that obtained using CCD-RSM optimization method. The ANN coupled with GA was found to be a powerful tool for optimizing production parameters with high level of accuracy. This technique may be used for other fermentation products to optimize the important process parameters before scaling up the process to industrial level.

Keywords: Process design, L-lysine, CCD-RSM, ANN-GA

Introduction

L-lysine is an essential amino acid with constant demand in the food and pharma industry. About 80% of the commercial production of lysine is achieved through microbial source whereas only 20% by the chemical means. The continuously growing demand of this amino acid requires cost compensation, due to the higher cost of carbon sources (e.g. glucose, fructose, sucrose, molasses), nitrogen sources (either inorganic and organic salts or both), nutrients such as potassium phosphate, magnesium sulphate, calcium chloride, zinc sulphate, sodium chloride also added if required for the industrial production of lysine[1-2]. Dietary intake of L-lysine has been shown to reduce chronic anxiety in humans. In communities where wheat is a staple diet, lysine-rich wheat is known to reduce anxiety and stress symptoms[3]. Also, it is proved to be effective in reducing the rigour and healing time for herpes simplex virus (HSV)[4] and has potential effect on schizophrenia[5]. Intake of L-lysine and L-arginine in combination reduces levels of salivary cortisol and hormonal stress responses, especially in human males[6]. L-lysine when consumed as a nutritional supplement like prolysin C and lysin C results in the reduction of skin patches in psoriasis patients[7]. According to a market report published by Transparency Market Research, the global L-lysine market revenue is expected to reach the US $7 billion by 2024 attributed to its increased usage in feed, pharmaceutical and cattle breeding sectors and their application will be much broader including feed supplement[8].

Application of mathematical and statistical tool for optimisation of substrate and production condition has become an integral part of the scale-up needs of the industry and many researchers have successfully used central composite design-response surface methodology (CCD-RSM) for optimisation and improved production of L-lysine[9]. These designs aim to determine the effect of individual parameters on overall production and to obtain a path to the optimum response which has been deriving through designing blocks and factorial experiments[10]. Production optimisation

*Authors for correspondence:
lalitdbtiitr@gmail.com, ashutoshcoolvns@gmail.com

and cost minimisation is an ever lasting quest in industrial production. Nutrient and environment have a colossal impact on the product from microbial cells. L-lysine is a growth associated primary metabolite. The optimisation of production and biomass along with nutritional balance is a matter of several combinatorial experimental runs, which is time taking and laborious process. Therefore, in this study L-lysine production has been used as a model for the optimisation and industrial production of amino acids, the validation of the hypothesis performed with central composite design (CCD) which is the most popular of the many classes of the RSM designs.

CCD can be run sequentially, is very efficient in providing much information on experimental variables effect and overall experimental error in a minimum number of required runs. Box and Wilson developed response surface methodology (RSM) with the objective of improving yield from various industries[11]. Presently, RSM has been applying in diverse fields like biotechnology, life science, process industries, automotive[12] and in conjunction with CCD it has been favourably using in the production of biomass, enzymes and several metabolites[13-15]. An alternative to RSM is the use of artificial neural network (ANN) modelling to obtain the desired accuracy level. ANN is a predictive tool comprising a set of algebraic equations and employs learning algorithms to decode the input output relationships between complexes, non-linear systems[16-17]. Genetic algorithm (GA) assisted by artificial neural network (ANN-GA) can used as a cogent tool for optimising process parameters and maximising L-lysine production. GA is based on natural selection and uses the three rules of selection, crossover and mutation to obtain optimal solution over successive generations.

Hypothesis testing and production optimisation was done with *Corynebacterium glutamicum* NCIM 2168. L-lysine production optimisation was studied using statistical tools like CCD-RSM, ANN-GA validation of CCD-ANN data. The parameters investigated for process optimisation were temperature, pH and glucose concentration. The models were used to determine the impact of the three process parameters on the concentration of L-lysine production used as a response for the evaluation of parameters. The developed models compared for their suitability for predicting L-lysine production, indicating the optimal approach.

Material and Methods

All chemicals used in this study were analytical grade procured from Sigma-Aldrich, Hi Media. The media and their ingredients were purchased from Hi-Media.

Microorganism

The strain of *C. glutamicum* NCIM 2168 obtained from the National Collection of Industrial Microorganisms (NCIM), NCL, Pune, India.

Growth Media

Nutrient agar media (Medium 41) was used to revive *C. glutamicum* NCIM 2168 (ATCC 13059) as per the composition provided by NCIM, NCL, Pune, India. The composition of media was (g/l): Beef extract 10.0; NaCl 5.0; peptone 10.0; agar 20.0 and pH adjusted between 7.0-7.5. The media was sterilized in an autoclave (121°C, 15 min).

Inoculum Preparation

The inoculum of *C. glutamicum* NCIM 2168 was prepared by transferring cells from agar slant into 2 ml growth media in test tubes. After incubation at 30°C, 120 rpm for 24 h, the freshly grown cells were transferred in the 10 ml growth media test tubes and incubated under above mentioned growth conditions. Finally 2% inoculum was used for production of L-lysine contained 10^6 cells per ml.

Production Media

The composition of L-lysine production media (per l) was as follows: 90 g D-glucose, 5 g $(NH_4)_2SO_4$, 8 g K_2HPO_4, 4 g KH_2PO_4, 0.2 g $MgSO_4.7H_2O$, 1.0 g NaCl, 0.5 g citric acid, 20 mg $FeSO_4.7H_2O$, 50 mg $CaCl_2.2H_2O$, 40 mg L-methionine, 100 mg L-leucine, 1 mg biotin, 1 mg thiamine HCl, trace salts in media (10 ml/L): 200 mg $MnSO_4$,100 mg $FeCl_3.6H_2O$, 1 mg $ZnSO_4.7H_2O$, 30 mg $CuSO_4.5H_2O$. The production run performed in 250 ml Erlenmeyer flask (100 ml culture volume) at 30°C and pH 7.0 at 120 rpm for analysing the growth kinetics of *C. glutamicum* NCIM 2168 and the confirmation of L-lysine production was done by ninhydrin analysis.

Process Optimisation

Experimental Design and Modelling

This study aimed to obtain the optimum values of process parameters namely, temperature, pH and glucose concentration for efficient and optimised production of L-lysine using CCD followed by RSM as statistical tool. Furthermore, ANN used as a prediction tool for process optimisation in order to further enhance productivity. A three-level-three

factor CCD was employed to conduct 20 experiments taking into consideration parameters such as temperature (°C), pH and glucose concentration (g/l). The actual and coded levels of the parameters are given in Table 1.

A 2^3 rotatable CCD followed by RSM was used to predict the influence of independent experimental factors and their interaction on maximum L-lysine production with different input values of the factors.

The experiments performed according to the CCD. Moreover, the obtained response through a CCD used for generating the best fit second order polynomial quadratic regression equation as given in equation 1:

$$Y= \beta_0 + \beta_1 A + \beta_2 B + \beta_3 C + \beta_{11} A^2 + \beta_{22} B^2 + \beta_{33} C^2 + \beta_{12} AB + \beta_{13} AC + \beta_{23} BC \quad \text{(Eq. 1)}$$

Where, Y is the dependent variable (L-lysine concentration, g/l), β_0 represents the offset value whereas β_1, β_2, β_3 are coefficients of linear terms β_{11}, β_{22}, β_{33} are quadratic coefficients and β_{12}, β_{13}, β_{23} denote interaction coefficients. A, B, C represents the independent variables, viz., temperature (°C), pH and glucose concentration (g/l), respectively. The effect of the parameters and their interaction on the response has been analysed by conducting significance tests and analysis of variance (ANOVA) on each response to check the adequacy of the model. Design Expert 9.0.6.2 (State-Ease Inc., USA) used for plotting the three-dimensional surface plots. The value of α calculated as 1.682, where $\alpha = 2^{k/4}$ (k = 3, the number of variables). The coded values of all independent variables and the experimental value of the only response variable Y (L-lysine concentration) along with predicted values presented in Table 2. The coefficients were calculated by Design Expert 9.0.6.2.

Table 1 — The range of parameters (independent variables) had chosen for the CCD.

Variables	Symbol	Coded levels		
		-1	0	+1
		Actual levels		
Temperature (°C)	A	27	30	32
pH	B	6.5	7	8.5
Glucose concentration (g/l)	C	70	95	120

Artificial Neural Network (ANN) Modelling

A multilayered feed-forward ANN with error back-propagation (BP) was employed using MATLAB R2016a (MathWorks Inc., USA) a strict learning scheme is of utmost significance for proper training of the network to assure useful mapping between inputs and outputs resulting in constant improvement of the network with effective error reduction[18-19]. A feed-forward network with back-propagation is most commonly used in process optimisation to minimise error at each iteration[20]. The developed ANN architecture was used to optimise L-lysine production using input neurons network topology. The number of neurons in hidden layer recognised by training of

Table 2 — Experimental plan, range and levels of independent (A), (B) and (C).

Std	Run	Factor 1 A: Temperature °C	Factor 2 B: pH	Factor 3 C: Glucose conc. g/l	Experimental L-lysine conc. g/l	Response CCD-RSM L-lysine conc. g/l	Response ANN L-lysine conc. g/l
13	1	29.5	7.25	52.96	9.68	15.68	8.246917
11	19	29.5	5.15	95	11.66	14.96	17.87883
3	11	27	8.5	70	13	10.1	15.54876
14	12	29.5	7.25	137.05	14.3	19.3	13.76141
4	14	32	8.5	70	14.66	14.66	14.20745
5	4	27	6	120	15	15.54	13.78808
9	10	25.30	7.25	95	15.3	12.3	15.54876
1	8	27	6	70	15.66	17.66	15.69821
2	9	32	6	70	16.33	16.33	15.77926
7	6	27	8.5	120	17	14.52	16.98033
6	3	32	6	120	17.33	17.33	16.79152
12	13	29.5	9.35	95	17.66	11.24	17.61684
10	2	33.71	7.25	95	18.66	17.66	18.51419
8	5	32	8.5	120	21.36	19.33	21.19988
18	20	29.5	7.25	95	26.46	17.93	28.18485
20	16	29.5	7.25	95	27.43	18.46	28.18485
15	17	29.5	7.25	95	28	18.67	28.18485
17	18	29.5	7.25	95	28	19.64	28.18485
19	7	29.5	7.25	95	28.36	19.86	28.18485
16	15	29.5	7.25	95	30.3	19.62	28.18485

several ANN topologies and selecting the optimal one which based on minimisation of mean square error (MSE) and overall correlation coefficient (R) to improve generalisation ability of the ANN topology. Experimental data obtained from CCD was used to construct and train neural network model. Overall 20 experimental data points were used out of which 70% used for training the network model while 30% (15% + 15%) for testing and validation of the model. This network model was trained based on the Levenberg-Marquardt (LM) back propagation algorithm in order to obtain the weights and biases. The neural network trained until the MSE reached a constant lower value with concomitant R-value (overall correlation coefficient) close to 1.

ANN-Network Design Optimization

In this study, the prediction of L-lysine production achieved with the help of multilayered feed-forward ANN, trained by back propagation[21]. A multilayered feed forward neural network, also known as multilayer perceptron (MLP) has been designed to function optimally while solving non-linear regression models[22]. An MLP consisted of the input layer, an output layer and one or more hidden layer with each layer having a specific number of neurons. The number of neurons in the input and output layers depends upon input variables and output variables respectively. The interconnection between neurons in each layer defined by weights and biases over which, signals transmit. Hidden layer (N_h) can lie between input (I) and 2I + 1 and that it should never be less than the maximum of input/3 and output. The determination of some hidden nodes always follows a trial and error approach[23]. Here, MLP architecture of ANN was used to build predictive model with three parameters temperature (°C), pH and glucose concentration (g/l) as input and experimental production of L-lysine (g/l) as output. This type of topology design also requires the specification of training algorithm, learning rate, number of iterations and retrains, and training stopping criteria. The input layer prepares scaled input data to be worked by the hidden layer through weights. These are some small, random, non-zero values trained by back propagation that ranged from −1 to +1. The activation functions transfer the weighted sum of the inputs to each hidden neuron as logistic sigmoid and then undergo another weighted sum transformation to get the outputs. Therefore, the number of neurons in the hidden layer possesses the capability to influence the accuracy of the network and hence is the most crucial criterion to be considered. The hidden layer sums up the weighted inputs along with biases as represented by the following equation 2

$$\text{Sum} = \sum_{i=1}^{n} x_i W_i + \theta \quad \text{(Eq. 2)}$$

Where, W_i (i =1, n) represents the weights of the connection between neurons of input and the hidden layer, θ defined as the bias and x_i represent the input parameter. The weighted output is transferred to a non-linear domain by an activation function which shifts the space in non-linearity of input data. The logistic function applied in the present study can be demonstrated by equation 3:

$$f(\text{sum}) = \frac{1}{1+\exp(-\text{sum})} \quad \text{(Eq. 3)}$$

The output thus produced by the hidden layer becomes an input to output layer, as neurons in the output layer produce output by neurons in the hidden layer. The calculated and actual experimental output has been formulated based on an error function. Training an ANN is an iterative process where this pre-specified error function is minimized by adjusting the weights appropriately. The commonly employed error functions mean square error (MSE) and root-mean-square error (RMSE) was used in this study equation 4

$$\text{MSE} = \frac{1}{N}\sum_{I=1}^{N}\left(\theta_{i,p} - \theta_{i,e}\right)\text{^}2 \quad \text{(Eq. 4)}$$

Where, N is the number of data points/experiments $\theta_{i,p}$ is predicted value obtained from the model and $\theta_{i,e}$ is the experimental value. In the present study, the efficiency of the model was decided based on the MSE, R^2, regression and correlation coefficients.

Genetic Algorithm

ANN combined with GA creates a potent tool for process modeling and optimisation for complex processes. GA solved complex optimisation problems and based on Darwin's principle of 'survival of the fittest'[24]. The algorithm initiates with a randomly selected set of chromosomes called a population. As the iteration (generation) proceeds, the process converges to stronger and filter solutions (chromosomes) obtained by reproduction among previous generation abiding by three key genetic operators which are selection, crossover and mutation. In the crossover, two best-fit parents are selected from existing generation and are combined to develop

offspring with good genes, and the process is iteratively continued through generations to converge to a good solution[25]. Offspring generated at each step evaluated for their fitness using developed ANN ensuring that crossover and mutation occur only among the best chromosomes. Point mutation is the most common form involved which creates minor changes in chromosomes with a predetermined probability[26].

L-lysine Production by Fermentation

Submerged fermentation was performed in 250 ml conical flasks with 100 ml of fermentation medium in each. The flasks were inoculated with *C. glutamicum* NCIM 2168 and placed in a shaking incubator (120 rpm). Five ml sample were taken at a set time interval (4 h) aseptically and centrifuged at 5000 rpm for 10 min. The supernatant used for analysis of L-lysine concentration and the biomass retained as a pellet. The pellet suspended in an appropriate amount of distilled water. Biomass growth was measured turbidometrically at 600 nm using calibration curve of *C. glutamicum* NCIM 2168.

The quantitative analysis of L-lysine was carried out by ninhydrin ferric reagent method[27]. Twenty µl of the sample mixed in 660 µl of reagent A (methylcellosolve 0.373 ml, 50% ferric chloride solution 30 ml and 0.1M KCl solution 600 ml) and 370 µl of reagent B (1% ninhydrin in 0.1M KCl solution). The solution was incubated in a boiling water bath (100˚C) for 20 min. The incubated sample was allowed to cool at room temperature. Four ml of dimethyl sulfoxide (DMSO) was added. The final volume was adjusted by addition of 3 ml distilled water before taking absorbance at 470 nm in a spectrophotometer. The concentration of produced L-lysine calculated from standard curve drawn by pure L-lysine.

Results and Discussion

L-lysine Production Optimisation

The experimental conditions for optimisation of L-lysine production was carried out according to the CCD in which temperature (A), pH (B), and glucose concentration (C) were used as variable parameters. The experimental runs are mentioned in Table 2, and its respective response (i.e. production of L-lysine in g/l) and the statistical analysis of experimental response by CCD is given in Table 3. This analysis was done to analyze the error produced during experiments before applying the RSM and ANN models for optimization, the outcome of this analysis were found as multiple R (0.975698308), R square (0.951987188) and standard error (1.937720129).

Optimisation by RSM Modeling

RSM simulated the experimental data for interaction analysis and their response plot. The L-lysine production ranged between 9.68 to 30.3 g/l. The analysis of variance (ANOVA) for L-lysine production using CCD-RSM is given in Table 4.

The model F-value of 26.93 implies the model is significant. Values of "Prob > F" less than 0.0500 indicated that the model terms are significant. In this case, A, B, C, AB, BC, A^2, B^2 are significant model terms. Values greater than 0.1000 indicate the model terms are not significant. The "lack of fit F-value" of 1.04 implied that the lack of fit is not significant relative to the pure error. Non-significant lack of fit is good to fit the model. The "Pred R-Squared" of

Table 3 — Statistical analysis of the experimental response of CCD.

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	9	744.485182	82.720576	22.03086	1.8906E-05
Residual	10	37.54759299	3.7547593		
Total	19	782.032775			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 99.0%	Upper 99.0%
Intercept	-603.4295522	95.46888358	-6.320694	8.673E-05	-816.1474809	-390.71162	-905.9964761	-300.8626284
A	30.68940672	5.183837509	5.92021	0.000147	19.13909697	42.239716	14.26041217	47.11840128
B	32.22455865	8.293133008	3.8856918	0.0030307	13.74630679	50.702811	5.941358842	58.50775846
C	1.06308878	0.392864708	2.7059921	0.022085	0.187731661	1.9384459	-0.182006603	2.308184163
A^2	-0.54176117	0.081668899	-6.633629	5.827E-05	-0.723730817	-0.3597915	-0.800592179	-0.282930161
B^2	-2.692018766	0.326678112	-8.240585	9.075E-06	-3.41990296	-1.9641346	-3.72735078	-1.656686753
C^2	-0.008240396	0.000816692	-10.08997	1.464E-06	-0.0100601	-0.0064207	-0.010828716	-0.005652076
AB	0.1208	0.219228007	0.5510245	0.5937133	-0.36767044	0.6092704	-0.573993331	0.815593331
BC	0.04144	0.021922801	1.8902694	0.0880229	-0.007407044	0.090287	-0.028039333	0.110919333
CA	0.00872	0.0109614	0.7955188	0.4447842	-0.015703522	0.0331435	-0.026019667	0.043459667

Table 4 — Analysis of variance (ANOVA) for L-lysine production using CCD-RSM.

Analysis of variance Table [partial sum of squares - type III]

Source	Sum of squares	df	Mean square	F value	p value (prob > F)	
Model	151.97	9	16.89	26.93	< 0.0001	significant
A-temperature	26.00	1	26.00	41.47	< 0.0001	
B-pH	15.41	1	15.41	24.57	0.0006	
C-glucose conc.	14.47	1	14.47	23.08	0.0007	
AB	9.92	1	9.92	15.83	0.0026	
AC	1.42	1	1.42	2.26	0.1633	
BC	13.03	1	13.03	20.78	0.0010	
A^2	23.56	1	23.56	37.57	0.0001	
B^2	54.42	1	54.42	86.79	< 0.0001	
C^2	2.20	1	2.20	3.52	0.0903	
Residual	6.27	10	0.63			
Lack of fit	3.20	5	0.64	1.04	0.4833	not significant
Pure error	3.07	5	0.61			
Cor total	158.24	19				

Std. Dev.	0.79	R-squared	0.9604
Mean	16.54	Adj R-Squared	0.9247
C.V. %	4.79	Pred R-Squared	0.8089
PRESS	30.24	Adeq Precision	17.284

0.8089 is in reasonable agreement with the "Adj R-Squared" of 0.9247; i.e. the difference is less than 0.2. "Adeq Precision" measures the signal to noise ratio. A ratio greater than 4 is desirable. The ratio of 17.284 indicated an adequate signal to navigate the design space.

Final Equation in Terms of Actual Factors

L-lysine = -120.82637 + 9.39719*0 temperature + 2.78976* pH-0.33484* glucose conc. + 0.35640* temperature * pH + 6.74000E-003* temperature * glucose conc. + 0.040840* pH * glucose conc. - 0.20457* temperature^2-1.24366* pH^2- 6.25799E-004* glucose conc.^{2} (Eq. 5)

The equation regarding actual factors can be used to make predictions about the response for given levels of each factor. By default, the high levels of the factors are coded as +1 and the low levels of the factors coded as -1. The actual factors equation (Eq. 5) is useful for identifying the relative impact of the factors by comparing the factor coefficients on solving the Eq. 5. The maximum amount of L-lysine was found to be 19.003 g/lysine.

The relation between the actual and predicted response presented in Figure 1. Interaction graph and contour plots showing the relationship between two parameters on keeping the remaining one as a constant.The interaction between pH and temperature shown in Figure 2a and 2b, data constant glucose concentration of 95 g/l. Through RSM, the optimum production of L-lysine was evaluated at pH 7.25 and temperature 29.5°C. Similarly, the interaction between glucose concentration and pH at a constant temperature of 29.5°C represented in Figure 3a and 3b and interaction between glucose concentration and temperature at constant pH 7.25 represented in Figure 4a and 4b.

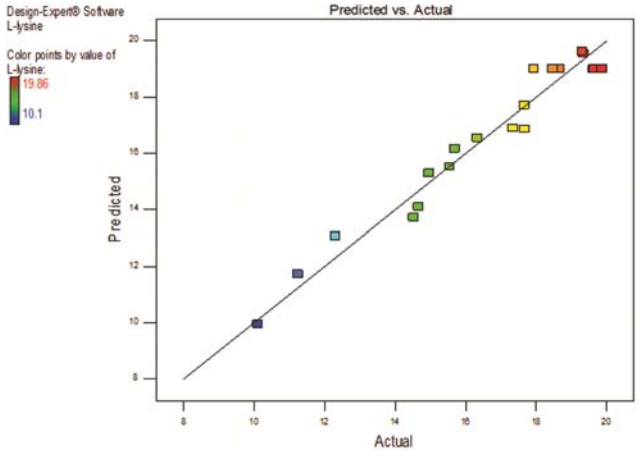

Fig. 1 — Relation between actual response and predicted response.

Optimization of the Number of Hidden Neurons

The neurons in the hidden layer had significantly affected accuracy and prediction of the optimal conditions. If the network topology is simple, the trained networks cannot learn properly. Therefore, neurons were optimised to determine optimum neuron using the best predictive capability and accuracy of the model. The neurons predicted were obtained based on model performance such as R^2, and MSE values. The optimum 20 neurons showed the best-predicted capability and high accuracy of the model

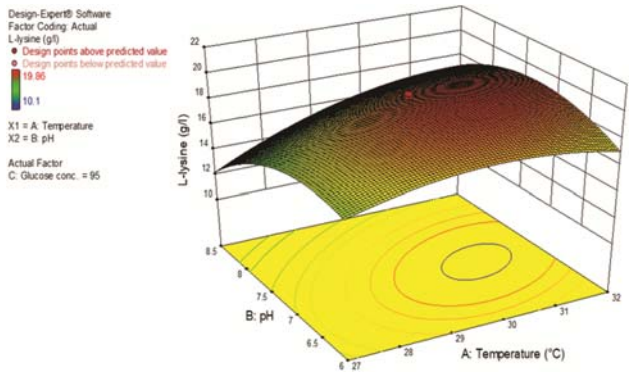

Fig. 2a — Interaction between pH and temperature.

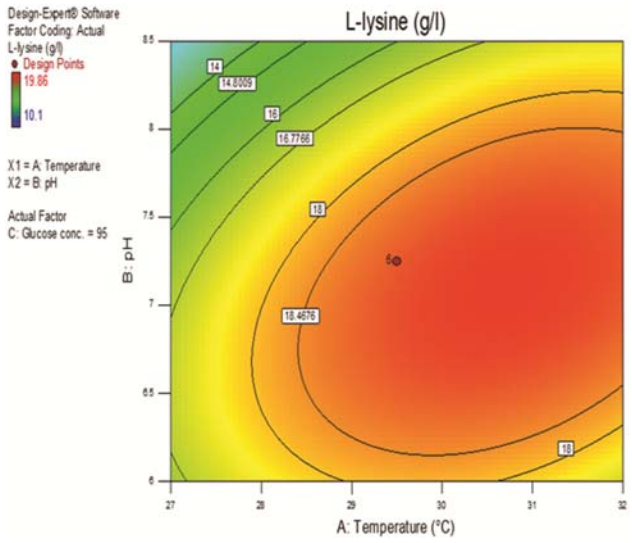

Fig. 2b — Contour graph interaction between pH and temperature.

for L-lysine production in Figure 5. The neural network models with the corresponding R^2 and MSE, mentioned in Table 5. Results showed the higher values of R^2 (0.9990), strongly suggests in a decrement of MSE values (0.397), which shows that developed ANN model was significant and could be used to predict optimal neuron for optimum production of L-lysine.

The MSE and R^2 values for the neural network with an optimal number of hidden neurons were found to be 0.397 and 0.9990, respectively. The active network topology for L-lysine production experiment was found to be 3-20-1, where 3 represent the input variables, 20 as hidden neurons and one as the output (Fig. 5). There are successful reports of the application of ANN-based optimisation. The optimal feed forward topology developed from neural network design used in the simulation of experimental outcome using ANN is mentioned in Table 2.

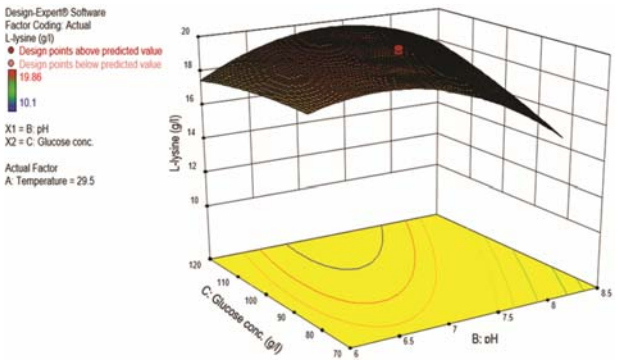

Fig. 3a — Interaction between glucose concentration and pH.

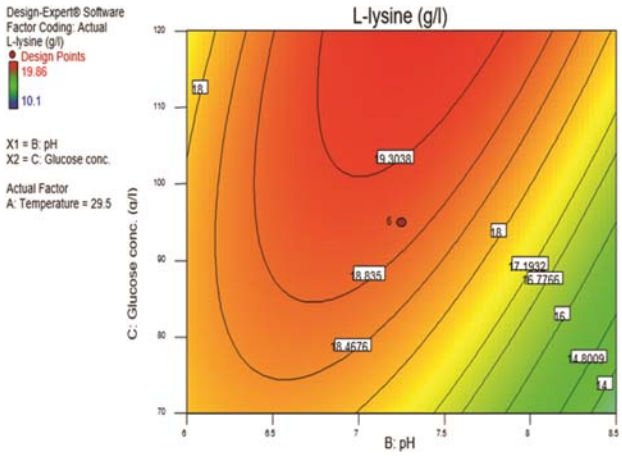

Fig. 3b — Contour graph interaction between glucose concentration and pH.

Table 5 — Neural network models with the corresponding R^2 and MSE.

ANN model design	Overall correlation coefficient (R^2)	Mean squared error (MSE)	Transfer function	
			Hidden layer	Output
3-10-1	0.9994	2.905		
3-12-1	0.9647	2.350		
3-14-1	0.9960	1.965	Tangent sigmoidal	Pure linear
3-16-1	0.9960	1.225		
3-18-1	0.9964	0.836		
3-20-1	0.9990	0.397		

Training, Validation and Testing of the Model

The input data divided into three sub-categories such as training (70%), validation (15%) and testing (15%) for the development of the model. Figure 7 represents the model of ANN with suitable R^2 values of training (0.9959), validation (0.9995) and testing (0.9952) and the overall model was best fit to a linear equation with R^2 value 0.9916 which was not close to 0.9604 (R^2 value of RSM data set). Thus, the developed ANN model was able to simulate accurately for L-lysine production (target) and reproduce experimental results with greater precision. Target had been precisely

achieved by the incorporation of multilayered feed forward ANN, trained by BP algorithm, with significant R^2 values. The quality of the data used to develop the ANN model is estimated by the error histogram plot. It was observed that most of the errors ranged between -0.6556 to 0.0187 (Fig. 8), However, the validation data point was observed with the highest error limit up to 0.2435 in comparison to the rest of the data set. Outliers are used to determine the quality of the given data. A large number of outliers in this model necessitated collection of more data points to improve the network.

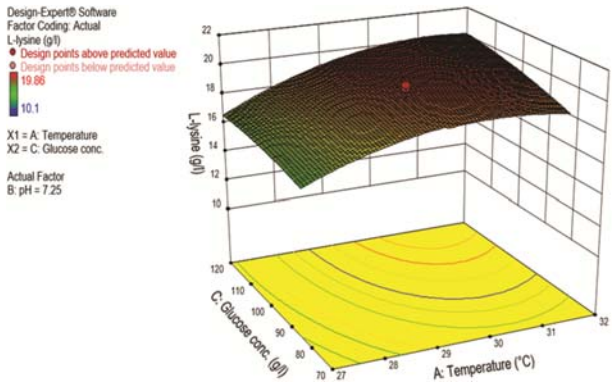

Fig. 4a — Interaction between glucose concentration and temperature.

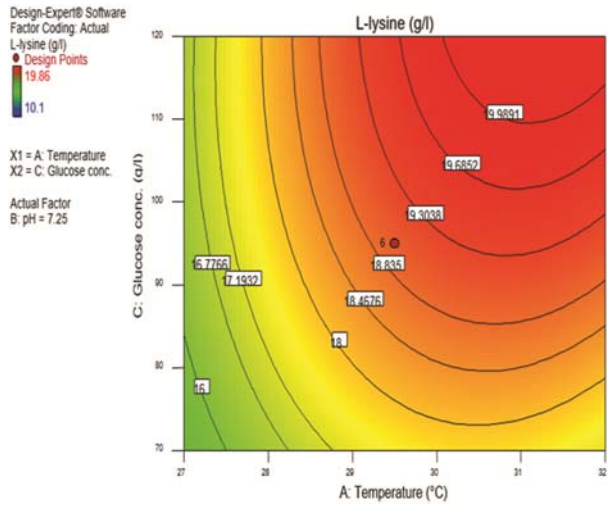

Fig.4b — Contour graph interaction between glucose concentration and temperature.

Comparison of RSM and ANN Model

A similar method was used by Das *et al* (2015) to compare the ANN and RSM data for validation. Figure 9 indicated that the data points of ANN response showing no overfitting and the model is statistically more suitable when compared to RSM response. The predictive capabilities of RSM and ANN models were compared on the basis of R^2 and MSE. The experimental and predicted values of L-lysine production by RSM and ANN showed in Table 2. The R^2 values for predicted models of RSM and ANN were found as 0.9604 and 0.9916, respectively. However, the MSE values for RSM and ANN were 0.61 and 0.397 respectively (Table 6). The higher predictive capability of ANN model attributed to its non-linear polynomials of the system whereas RSM can generalise data by only quadratic equations. The comparative predictive supremacy of ANN over experimental response has been reported by some researcher[28].

The use of RSM or ANN mainly depends upon the data set type. However, ANN has been known to

Table 6 — Comparison of the predictive capacity of RSM and ANN.

Parameters	RSM	ANN
R square (R^2)	0.9604	0.9916
Mean square error (MSE)	0.61	0.397

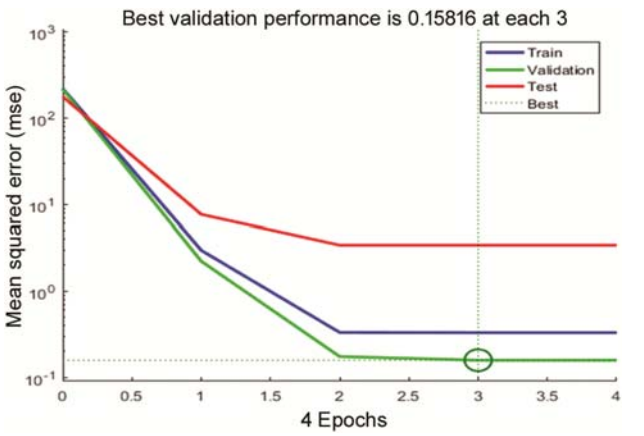

Fig. 6 — Performance plot during the ANN training.

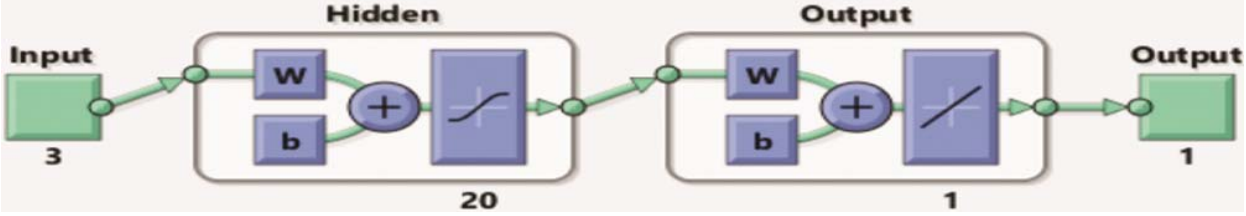

Fig. 5 — ANN architecture topology with input, hidden (tangent sigmoid transfer function) and an output layer (pure linear transfer function).

show a higher predictive accuracy as compared to RSM. The RSM based models are structured in nature and are useful for obtaining sensitivity analysis and relationship between different input and output components when they are present in a limited number[29]. RSM models are however not entirely feasible for highly nonlinear processes. The ANN, on the other part, can learn and generalise the behaviour of any complex process and represents non-linearities in a much better way than RSM[30].

Validation of the ANN Model by GA

GA used as an optimisation tool for the developed ANN model. The GA tool kit of MATLAB 2016a (MathWorks Inc., Natick, USA) used in this study for GA analysis. The L-lysine production, which is dependent variable, fixed as a chromosome. Each chromosome consisted of 3 genes, i.e. independent

Table 7 — Parameters for validating the ANN response by GA (ANN-GA)

Operation	Input parameters
Solver	Genetic algorithm
Variable	3
Fitness function	function Y= optim (X) Y= -[-607.7180+32.4635*X(1)+27.7932*X(2)+1.011268198 *X(3)-0.5710*X(1)^2-2.1223*X(2)^2-0.00912056 *X(3)^2-0.00825 *X(1)*X(2)+0.03728748 *X(2)*X(3)+0.0169665 *X(3)*X(1)];end
Bounds	Lower [25 5 70] Upper [32 8 120]
Population	Double vector
Creation function	Feasible population
Selection	Tournament
Reproduction	0.8 (default)
Mutation	Point mutation
Crossover function	Constrain dependent
Optimum	Temp- 29.83 pH-7.35 Glucose conc. 98.21

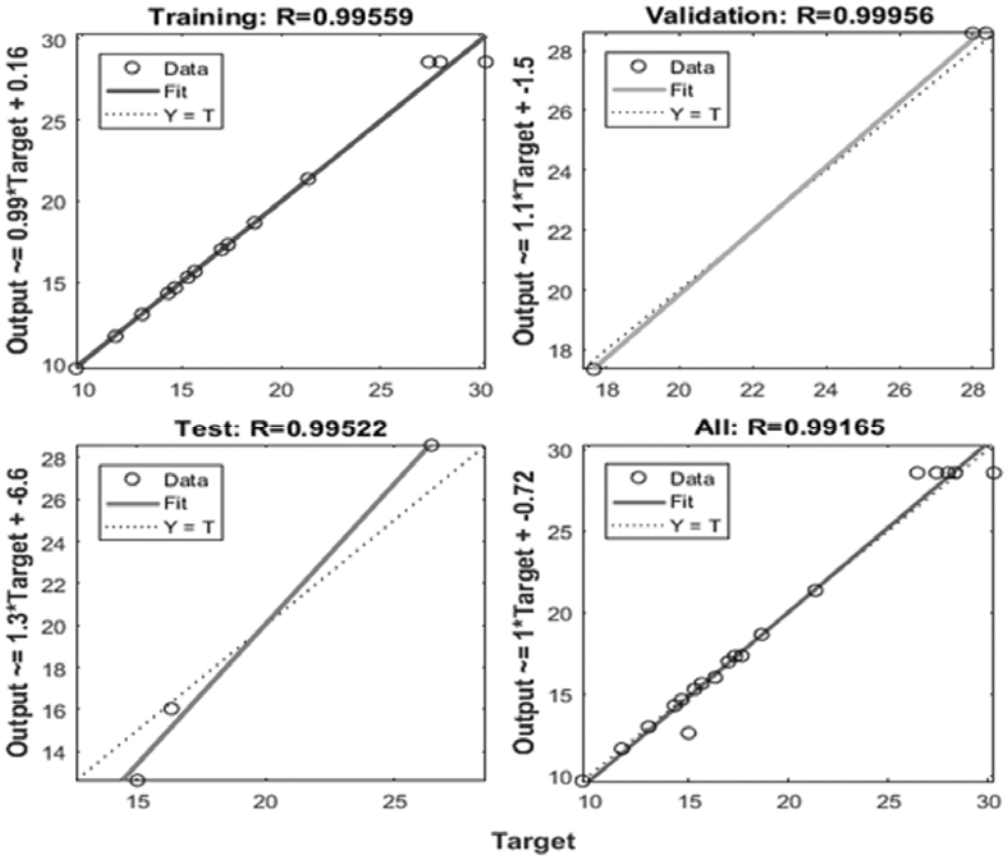

Fig. 7 — Regression of experimental and predicted values using ANN.

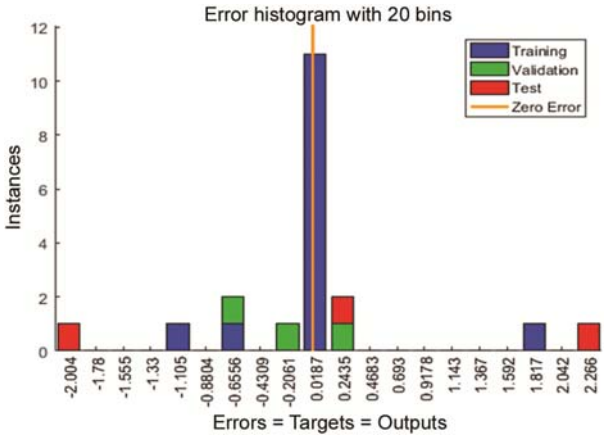

Fig 8 — Error histogram plot for the ANN model for L-lysine production.

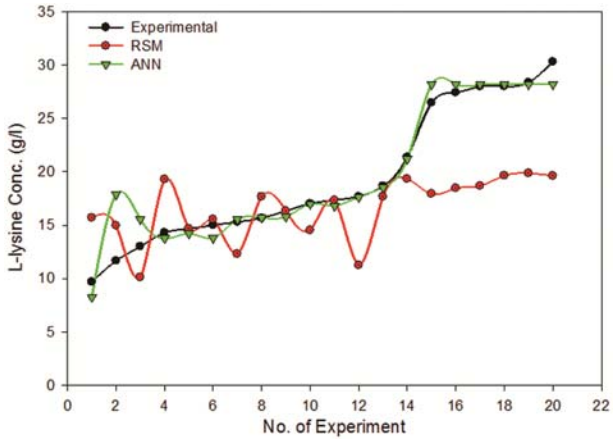

Fig. 9 — Comparison chart of experimental vs RSM vs ANN data.

variables such as temperature, pH and glucose concentration. Parameters for validating the ANN response by genetic algorithm (ANN-GA) to screen the optimum values of each gene are mentioned in Table 7. Fitness value of ANN-GA represents the close relationship between best fitness and mean fitness is -28.363.

External Validation of GA Parameters

The predicted condition for optimized production of L-lysine was pH 7.35, temperature 29.83°C and glucose concentration 98.21 g/l. The predicted condition was further validated experimentally in triplicates, and the L-lysine production was achieved up to 27.25 ± 1.15 g/l. The predicted value with ANN-GA was close enough to the experimental value.

Conclusion

This study designed as a model for production optimization of primary metabolite for industrial application. The ANN-GA coupled model found to be better tool for optimization of L-lysine production. The predicted production of L-lysine through ANN-GA was 28.363 g/l at temperature 29.83°C, pH 7.35 and glucose concentration 98.21 g/l. However, the actual run had limitation of incubation temperature therefore, 30°C used instead 29.83°C, at this condition the actual production was 27.25 ± 1.15 g/l. The ANN-GA optimization was found to be better than the CCD-RSM (19.003 g/l) based on experimental data and validation. It is recommended to use of ANN-GA as appropriate and alternative method for production optimization.

Acknowledgements

Authors are thankful to Department of Biochemical Engineering, Harcourt Butler Technical University, Kanpur, India for providing the required infrastructure, Authors acknowledge the CEES, INDIA for providing necessary guidance.

References

1 Coello N, Brito L & Nonus M, Biosynthesis of L-lysine by *Corynebacterium glutamicum* grown on fish silage, *Bioresour Technol*, 73 (2000) 221-5.

2 Bisaria R, Madan M, & Vasudevan P, Utilisation of agro-residues as animal feed through bioconversion, *Bioresour Technol*, 59 (1997) 5-8.

3 Smriga M, Ghosh S, Mouneimne Y, Pellett P L & Scrimshaw N S, Lysine fortification reduces anxiety and lessens stress in family members in economically weak communities in Northwest Syria, *Proc Nat Acad Sci*, 101 (2004) 8285-8288.

4 Griffith R S, Walsh D E, Myrmel K H, Thompson R W & Behforooz A, Success of L-lysine therapy in frequently recurrent herpes simplex infection, *Dermatology*, 175 (1987) 183-190.

5 Wass C, Klamer D, Katsarogiannis E, Pålsson E, Svensson L *et al*, L-lysine as adjunctive treatment in patients with schizophrenia: A single-blinded, randomized, cross-over pilot study, *BMC Medicine*, 9 (2011) 40.

6 Smriga M, Ando T, Akutsu M, Furukawa Y, Miwa K *et al*, Oral treatment with L-lysine and L-arginine reduces anxiety and basal cortisol levels in healthy humans. *Biomedical Research*, 28 (2007) 85-90.

7 Wong A P, Kalinovsky T, Niedzwiecki A & Rath M, Efficacy of nutritional treatment in patients with psoriasis: A case report, *Exp Ther Med*, 10 (2015) 1071-1073.

8 Lysine and other amino acids (methionine, threonine & amp; tryptophan) market- global industry analysis and forecast 2018 [Internet]. Food & Beverages. 2013. p. 100. Available from: https://www.transparencymarketresearch.com/lysine-amino-acids.html

9 Coello N, Montiel E, Concepcion M & Christen P, Optimisation of a culture medium containing fish silage for L-lysine production by *Corynebacterium glutamicum*, *Bioresour Technol*, 85 (2002) 207-211.

10 Box G, Hunter W C & Stuart J C, *Statistics for experimenters. an introduction to design, data analysis, and*

model building (No. 519.5 B69s Ej. 1) John Wiley & Sons, (1978).

11 Box G E, The exploration and exploitation of response surfaces: Some general considerations and examples. *Biometrics*, 10 (1954), 16-60.

12 Myers R H, Montgomery, D C Vining G G, Borror C M & Kowalski S M (2004), Response surface methodology: A retrospective and literature survey, *J Qua Technol*, 36 (2004), 53-77.

13 Udeh K O & Achremowicz B, Optimization of cultivation medium composition of an L-lysine producing mutant: The use of response surface methodology, *Acta Microbiologica Polonica (Poland)*, (1993).

14 Saval S, Pablos L & Sanchez S, Optimization of a culture medium for streptomycin production using response-surface methodology, *Bioresour Technol*, 43 (1993), 19-25.

15 Ergun M & Mutlu S F, Application of a statistical technique to the production of ethanol from sugar beet molasses by *Saccharomyces cerevisiae*, *Bioresour Technology*, 73 (2000), 251-255.

16 Vaidya R, Vyas P & Chhatpar H S, Statistical optimization of medium components for the production of chitinase by *Alcaligenes xylosoxydans*, *Enzyme Microbial Technol*, 33 (2003), 92-96.

17 Zobel C W & Cook D F, Evaluation of neural network variable influence measures for process control, *Eng Appl Artifi Intell*, 24 (2011), 803-812.

18 Beale M H, Hagan M T & Demuth H B, Neural network toolbox user's guide, The MathWorks *Inc, Natick, MA*, (2010).

19 Witek-Krowiak A, Chojnacka K, Podstawczyk D, Dawiec A & Pokomeda K, Application of response surface methodology and artificial neural network methods in modelling and optimization of biosorption process, *Bioresour Technol*, 160 (2014), 150-160.

20 Turan N G, Mesci B & Ozgonenel O, The use of artificial neural networks (ANN) for modeling of adsorption of Cu (II) from industrial leachate by pumice, *Chem Eng J*, 171 (2011), 1091-1097.

21 Gazzaz N M, Yusoff M K, Aris A Z, Juahir H, & Ramli M F, Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors, *Mari Poll Bull*, 64 (2012), 2409-2420.

22 Das S, Bhattacharya A, Haldar S, Ganguly A, Gu S *et al*, Optimization of enzymatic saccharification of water hyacinth biomass for bioethanol: Comparison between artificial neural network and response surface methodology, *Sustain Mat Technol*, 3 (2015) 17-28.

23 Singh K P, Basant A, Malik A & Jain G, Artificial neural network modeling of the river water quality: A case study, *Ecol Model*, 220 (2009) 888-895.

24 Goldberg D E & Holland J H, Genetic algorithms and machine learning, *Machine Learning*, 3(1988), 95-99.

25 Gupta A K, Guntuku S C, Desu R K, & Balu, A (2015), Optimisation of turning parameters by integrating genetic algorithm with support vector regression and artificial neural networks, *Int J Adv Manuf Technol*, 77 (2015) 331-339.

26 Singh A, Majumder A & Goyal A, Artificial intelligence based optimization of exocellular glucansucrase production from *Leuconostoc dextranicum* NRRL B-1146, *Bioresour Technol*, 99 (2008), 8201-8206.

27 Hsieh C L, Hsiung K P & Su J C, Determination of lysine with ninhydrin-ferric reagent, *Anal Biochem*, 224 (1995) 187-9.

28 Desai K M, Survase S A, Saudagar P S, Lele S S & Singhal R, Comparison of artificial neural network (ANN) and response surface methodology (RSM) in fermentation media optimization: Case study of fermentative production of scleroglucan, *Biochem Eng J*, 41 (2008), 266-273.

29 Bezerra MA, Santelli RE, Oliveira EP, Villar L S & Escaleira L A, Response surface methodology (RSM) as a tool for optimization in analytical chemistry, *Talanta*, 76 (2008) 965-77.

30 Hajmeer M N, Basheer I A & Najjar Y M, Computational neural networks for predictive microbiology II - Application to microbial growth, *Int J Food Microbiol*, 34 (1997), 51-66.