International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 8, Issue 7, 2016

Original Article

BIOPROCESS MODELING FOR THE PREDICTION OF THERAPEUTIC ENZYME L-ASPARAGINASE ACTIVITY IN SOLID STATE FERMENTATION USING MULTIPLE LINEAR REGRESSION AND ANN

ARCHANA VIMAL, JUJJAVARAPU SATYA ESWARI, AWANISH KUMAR*

Department of Biotechnology, National Institute of Technology (NIT), Raipur 492010, India
Email: drawanishkr@gmail.com

Received: 08 Apr 2016 Revised and Accepted: 20 May 2016

ABSTRACT

Objective: L-asparaginase is an enzyme of industrial as well as therapeutic importance. The capabilities of bioprocess modeling of L-Asparaginase activity produced from *Aspergillus niger* by solid state fermentation (SSF) were explored here.

Methods: Regression modeling (RM) and Artificial Neural Network (ANN) techniques were applied on input process parameter, which includes solid substrate, temperature, moisture percentage, particle size, cooking time to optimize L-Asparaginase enzyme activity in SSF.

Results: The L-asparaginase activity were obtained 38.918 (U/gds) and 38.714 (U/gds) with the optimum input parameters (x_1= *Glycine max*, x_2=30 (°C), x_3=6.5, x_4=70 (%), x_5=1180(µ), x_6=30 min) by ANN, and (x_1= 3, x_2=30 (°C), x_3=6.5, x_4=70 (%), x_5=1305(µ), x_6=30 min) by RM respectively. The goodness of fit of the model was determined in terms of R^2. The value of R^2 obtained by ANN after training and validation and over all data was 0.996, 0.989 and 0.981, whereas the value of R^2 obtained with linear, quadratic and full regression models was 0.501, 0.910 and 0.914 respectively.

Conclusion: This hybrid ANN/RM effectively identifies the significant process parameters and optimum production of L-asparaginase in the given larger set of conditions and able to reduce the number of experiments. Optimization by these modeling methods predicts the good activity of the enzyme and indicating its suitability and applicability for bioprocess modeling.

Keywords: L-asparaginase, Solid state fermentation, Regression modeling, Artificial neural network, Activity

INTRODUCTION

L-Asparaginase (EC.3.5.1.1; asparagine amidohydrolase), is an enzyme of chemotherapeutic importance. It is used in the treatment of acute lymphoblastic leukemia, lymphosarcoma, acute myeloblastic leukemia, chronic lymphocytic leukemia, Hodgkin's disease, melanosarcoma, and non-Hodgkin's Lymphoma. Normal cells have L-asparaginase synthetase, which produces L-asparagine for their growth and multiplication, whereas cancerous cell lacks L-asparaginase synthetase. Therefore, they are dependent on the L-asparagine of plasma pool for their proliferation. L-asparaginase deaminates asparagine to asparatic acid with the liberation of ammonia. It ultimately causes the death of cancerous cell due to nutritional starvation [1]. This enzyme may be produced by various sources like plants, animals, bacteria, and fungus. But commercially they are produced from *Escherichia coli* and *Erwinia chrysanthemi*are [2]. Due to high medical importance and huge market demand of L-asparaginase, industries are in constant search of better producing strains, and methods of its production like solid state fermentation (SSF). L-asparaginase is produced by submerged fermentation at industrial scale but researchers are looking forward for SSF at industrial scale production because SSF is an economic process that gives high yield at low capital investment [3]. SSF is also an environment friendly method because it produces less waste water and reduces the problem of disposal of solid waste generated [4].

We have used Regression modeling (RM) and Artificial Neural Network (ANN) for accurate prediction of L-asparaginase activity in SSF. These analysis were done to standardize data interpretation into a uniform and reproducible way. It depicts the experimental data as a mathematical equation in the form of $y=f(x)$, where x is the 'independent' variable and is controlled by the experimenter; y is the 'dependent' variable, which is measured; and f is the function, which includes one or more parameters used to describe the data. Better the fit, more accurately the function can be described [5]. Modeling of experimental data using ANN is a decent approach to perform estimations, control studies, fault diagnosis, and classification analysis. ANN is widely used as a model for the study of complex and non-linear behavior of the systems with a reliable set of input and output data pairs without the need of any prior information [6]. ANN would lead to save the time and cost by predicting the results of the reactions, so that the most promising conditions can be verified easily [7]. Production of L-asparaginase activity is greatly influenced by fermentation process, media, and culture conditions such as temperature, pH, inoculum size, moisture, particle size, agitation rate, and incubation time. The selection of appropriate substrate is a crucial factor in SSF in terms of nutritional requirements of growing microorganism as well as low water activity and high oxygen transference. Although many researchers have reported the optimization of carbon and nitrogen sources for L-asparaginase production but very less research work is found on the influence of culture conditions [8].

This serves the basis of modeling study on the report published by for predicting L-asparaginase activity [9]. In this paper, by Abha Mishra, 2006 describes the L-asparaginase production from *Aspergillus niger* by SSF employing three leguminous crops (*Cajanus cajan, Phaseolus mungo,* and *Glycine max*). The RM and ANN are performed to obtain a measure of the degree of association or correlation existing between the enzyme activity (output) and various experimental input parameters (media, temperature, pH, moisture, particle size, and cooking time). They are significant parameters for the optimization of L-asparaginase production and RM & ANN are good tool for the prediction of enzyme activity.

MATERIALS AND METHODS

Experimental data collection

The data (Supplementary table 1) used in this study was collected from the paper entitled "Production of L-Asparaginase, an anticancer agent, from *Aspergillus niger* using agricultural waste in solid state fermentation" by Mishra, 2006. There were six input parameters considered during this study that are temperature, pH, moisture, particle size, and cooking time. The fermentation output,

L-Asparaginase activity, was modeled by regression analysis using the 67 data points. One enzyme unit is defined as the amount of the enzyme, which liberates 1 µM of ammonia per gram of dry substrate per minute (U/gds) under optimal assay conditions.

Regression modeling

SSF data of *A. niger* were modeled by regression analysis using Microsoft Excel (2007) to estimate L-asparaginase activity. Media, temperature, pH, moisture, the particle size of the solid substrate, and cooking time were used as the input. Enzyme activity was output which is to be optimized by RM and ANN.

Regression analysis of only linear terms

In the initiation of modeling only linear terms were considered which were designated as x_1= media, x_2= temperature (°C), x_3= pH, x_4= moisture (%), x_5= particle size (µ), x_6= cooking time (minute) respectively.

Regression analysis of linear and quadratic terms

Squaring terms of input parameters (media, temperature, and pH) were also used for regression modeling further to increase the value of R^2 along with their linear value.

Regression analysis of combinations of linear, quadratic, and interaction terms

The final step in regression modeling was to incorporate linear interaction terms along with linear and quadratic terms to obtain a better value of R^2 to establish a greater correlation between input and output parameters [10, 11].

Artificial neural network methodology

ANN consists of 3-layers with the feed-forward network (MFFN) with multi-layered perceptron (MLP). A mapping between an input (x) and output (y) through a nonlinear function *f, i.e.*, $y = f(x)$ is called MFFN. The three layers of neural networks are input nodes, hidden nodes, and output layers nodes. Weighted links are called connecting links since each and every node of the input layer are connected to every node of hidden. Along with 3 nodes, there is another node called bias node is processing constant represented by the output of 1. All the numerical operations are operated through hidden and output layer nodes are hence known as active nodes, but input layer will not involve in numerical operation.

Training algorithm

The major complication in training the neural network is to minimize the prediction error which can be achieved by obtaining a set of weights so as the difference between predicted outputs by the network and the desired outputs can be minimized. For the prediction of the new input condition, the network identifies the patterns in data by the process of iterative training and generates an internal model. There are n-dimensional vector x_p, and a unit bias as input to the network. To have the activation state S_{pi}, in each input weight w_{ij} is multiplied and the products are summed which is given below in equation 1.

$$S_{pj} = \sum_{i=1}^{N} w_{ij} x_{pi} + w_{N+1,\,j} \quad \text{..... (1)}$$

The output of the hidden layer neuron O_{pj} for sigmoid function is calculated as:

$$O_{pj} = f(S_{pj}) = \frac{1}{1+e^{-S_{pj}}} \quad \text{..... (2)}$$

Where *f* represents the differentiable and non-decreasing function. Same calculation as mentioned above was performed by output layer of a single hidden layer network except that the input vector x_p is replaced by the hidden layer output O_p and the corresponding weights w_{jk}:

$$S_{pk} = \sum_{i=1}^{M} w_{jk} O_{pi} + w_{M+1,k} \quad \text{..... (3)}$$

$$O_{pk} = y_{pk} = \frac{1}{1+e^{-S_{pk}}} \quad \text{..... (4)}$$

A network containing more than one hidden layer may utilize similar calculations. A basic method that defines sum of squared error, E_p for p learning patterns was used to monitor the progress of learning. There are p input-output vector pairs (x_p, d_p) as a set of training examples. Weights are randomized initially and then to minimize the objective function $E(w)$ weights are adjusted. The mean squared error between the prediction outputs is defined as $E(w)$. The y_{pk} and d_{pk} for all the input patterns:

$$E(w) = \sum_{p=1}^{p} E_p \quad \text{..... (5)}$$

Where E_p is the sum of squared error with each training example,

$$E_P = \sum_{K=1}^{M} (d_{pk} - y_{pk})^2 \quad \text{..... (6)}$$

Gradient descent technique such as generalized delta rule is applied for training the network, so as the E_p minimization is achieved. This rule states that the error function δ_{pk} between the hidden layer neurons to the output layer neuron k is calculated as:

$$\delta_{pk} = (d_{pk} - y_{pk}) f'(S_{pk}) \quad \text{..... (7)}$$

δ_{pj} that is the error function from input neuron to hidden neuron j can be calculated as

$$\delta_{pj} = f'(S_{pj}) \sum_{k=1}^{M} \delta_{pk} w_{jk} \quad \text{..... (8)}$$

The change in weight from output to hidden layer after n^{th} data presentation is given by

$$\Delta w_{jk}(n) = \eta \delta_{pk} O_{pk} + \alpha \Delta w_{jk}(n-1) \quad \text{..... (9)}$$

Where η the learning is rate and α is the momentum factor. The updated weights are given by

$$w_{jk}(n) = w_{jk}(n-1) + \Delta w_{jk}(n) \quad \text{..... (10)}$$

To calculate the weight change from hidden layer to the input layer, a similar method was opted. Then randomly a new training example was selected when the weights were updated, until and unless satisfactory reduction of the objective function was obtained [13].

Information processing

The network training is an iterative process that begins with random initialization of weight matrices. In network learning process there are two types of passes that are forward pass and reverse pass. In forward pass network learning process, an input pattern from the example data set of the input nodes is applied. Then the calculation of the weighted sum of the inputs to the active node is performed, followed by transformation into output by using a nonlinear activation function such as sigmoid function.

In this way, the outputs of the hidden nodes (that is computed from the inputs to the output layer nodes) are evaluated in the similar fashion. On the other hand, the pattern specific squared error defined previously in the equation (5) is computed in the reverse pass. Then updation of network weights according to the gradient strategy was performed using the computed value in the previous step. Weight updating procedure is repeated for all the patterns in the training dataset that account for the completion of one iteration. The number of nodes in input and output layer in any modeling problem is based on ANN and governed by the input-output dimensionality. But the number of hidden nodes can be adjusted in order to prevent the

oversize of the network. By conducting network simulations the number of 'hidden units,' network over-fitting can be avoided. These simulations not only impart optimal network architecture but also reduce error to the minimum magnitude of the test data.

Training and testing procedure

The whole data were split into two data sets. One set is bigger that consist of 57 data which were used for training. The rest 10 data were incorporated in the smaller set; that was kept aside for further testing and validation of the ANN predicted output values. All the input values must be in the range of 0 to 1. Therefore, first of all input values were normalized. After normalization following steps were performed for training the network.

Step 0: Initialize the weight by setting random values in between 0 to 1.

Step 1: When stopping condition is false or not satisfied, perform steps 2-9.

Step 2: For every training pair of set steps 3-8 is to be performed

Step 3: Each input unit (x_i, i = 1, 2,. , 6) acquires input signals x_i and transfers this signal to every nodes in the next hidden layer.

Step 4(a): Each and every hidden unit (h_j, j = 1, 2,. ., 15) sum up its weighted input signals and then the bias is added to this so that output signal O_{pj} is calculated and this signal is forwarded to every unit in the next hidden layer h_k.

Step 4(b): Each hidden unit (h_k, k = 1, 2,. ., 16) sum up its weighted input signals and then bias is added to this so that output signal O_{pk} is calculated. This signal is forwarded to every unit in the next layer, and that is output layer.

Step 5: In the output layer weighted input signals are summed up by the output unit, and then to compute its output as mention in equation 4 applies its activation function.

Step 6: At output layer the BP of errors starts so as to compute its error information term. The BP of errors starts at the output layer to compute its error information term. It calculates weights correction and bias correction terms in equation 7 and 8 respectively.

Step 7(a): From units in the next layer delta inputs of each hidden unit (h_k, k = 1, 2,. ., 16) is summed up so as to compute its error information term and also calculates its weights correction and bias correction term.

Step 7(b): As step 7(a), from units in the next layer delta inputs of each hidden unit (h_k, k = 1, 2,. ., 15) is sum up so as to compute its error information term and also calculates its weights correction and bias correction term.

Step 8: Each output node (O_l, l = 1) updates its weights and bias given in equation 9 and 10.

Step 9: Stopping condition is tested.

To upgrade the back propagation algorithm, the Levenberg-Marquardt variation of nonlinear least squares optimization technique was employed. This technique was utilized to enhance the learning rate and the BP algorithm for searching minimum error. The algorithm and data processing were performed by using the MATLAB-2012 package.

RESULTS AND DISCUSSION

Regression modeling results

The coefficient of determination was calculated, which measures the strength of the relationship that exists between the input variables and output (L-asparaginase activity). More the value of R^2 towards 1 suggests more accurately functions fit the data. The value of R^2 obtained was 0.501, when only linear terms were considered that is not a good fit is. The equation 11 representing the output in terms of y_1 developed on the basis of linear parameter analysis is mentioned below which is not very reliable as the value of R^2 is very low.

$$y_1 = 3.9273 + 4.1256x_1 - 0.1097x_2 - 0.0187x_3 + 0.2560x_4 + 0.0005x_5 - 0.1475x_6 \ldots\ldots (11)$$

Where x_1= media, x_2= temperature (°C), x_3= pH, x_4= moisture (%), x_5= particle size (µ), x_6= cooking time (minute)

The graphical comparison of the experimental and calculated output of regression analysis is shown in fig. 1 (a). The difference between experimental and calculated output was high, which further suggested for the consideration of quadratic terms. Therefore, we again go for regression analysis after squaring the terms. After adding the squaring terms, the value of R^2 obtained was 0.910, which is a good fit. y_2 is the output shown in the equation developed on the basis regression modeling after adding quadratic terms mentioned below (equation 12) which is quite reliable as the value of R^2 is near to 1.

$$y_2 = -2228.225 - 6.6150x_1 + 3.5592x_2 + 35.9106x_3 + 1.0262x_4 + 0.0367x_5 + 0.9854x_6 + 2.8786x_1^2 - 0.0544x_2^2 - 2.8408x_3^2 - 0.0073x_4^2 - 9.6x_5^2 + 0.015110^{-06}x_6^2 \ldots\ldots\ldots (12)$$

Comparison graph of the experimental and calculated output of regression analysis is shown in fig. 2 (b). It is observed that the difference between the experimental and calculated output of regression analysis is minimized in this case. But for further improving the value of R^2 we proceed for regression analysis using linear, squaring and interaction parameter. Equation 13 is representing the output in terms of y_3, which is developed after considering linear terms, quadratic terms, and the interaction terms that can be used for the prediction of enzyme activity. In this case, the value of R^2 is 0.914.

Hence, the polynomial model given in equation 13 is good for prediction of experimental results. Reliability of the model is predicted on the basis of values obtained after regression analysis that is summarized in table 1.

$$y_3 = 2.6927 + 0.3253x_1 - 4.5805x_2 + (0 \times x_3) + 1.0327x_4 + 0.0370x_5 + 0.9896x_6 + 2.8828x_1^2 - 0.0546x_2^2 - 2.9301x_3^2 - 0.0074x_4^2 - 9.7 \times 10^{-06}x_5^2 + 0.0152x_6^2 - 0.0891x_1x_2 - 0.6410x_1x_3 + 1.2819x_2x_3 \ldots\ldots\ldots (13)$$

Table 1: Summary output in the optimization of operating conditions

Regression statistics	
Multiple R	0.956351
R^2	0.914607
Adjusted R^2	0.872386
Standard Error	2.438671
Observations	67

Fig. 1 (c) shows the difference between experimental output and output calculated from the derived formula. Table 2 represents the ANOVA (Analysis of Variance) summary of the full regression model, on the basis of which statistical significance of the model can be predicted. ANOVA is used to test the model and goodness of fit. The 'significance F' indicates the probability that the regression output could have been obtained by chance. The value of 'significance F' should be less than 0.05 for accepting the result of regression modeling. In this case, the value of 'significance F' is very less (7.623 x 10-[23]), which further indicates the model is highly significant with a meaningful correlation. For all (linear, quadratic, and the interaction terms considered together) coefficients of the operating conditions in the optimization of operating conditions is given in table 3. Table 3 also includes the t-value and p-value at 95 % significance which means that the p-value should be less than 0.05 for the acceptance of this model.

Table 2: Result of ANOVA test

	Df	SS	MS	F	Significance
Regression	15	3312.24	220.816	39.7821	7.62313E-23
Residual	52	309.25	5.94712		
Total	67	3621.49			

Table 3: Estimated regression coefficients in optimization of operating conditions

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	2.6927716	21.2801	0.12654	0.89979	-40.009	45.3945	-40.009	45.3945
x_1	0.32537332	5.27445	0.06169	0.95105	-10.259	10.9093	-10.259	10.9093
x_2	-4.5805347	1.22004	-3.7544	0.00044	-7.0287	-2.1323	-7.0287	-2.1323
x_3	0	0	65535	#NUM!	0	0	0	0
x_4	1.03276892	0.11389	9.06826	2.7E-12	0.80424	1.2613	0.80424	1.2613
x_5	0.03703905	0.00517	7.16499	2.7E-09	0.02667	0.04741	0.02667	0.04741
x_6	0.98961818	0.19788	5.00098	6.9E-06	0.59253	1.3867	0.59253	1.3867
x_1^2	2.88285725	0.65793	4.3817	5.7E-05	1.56262	4.20309	1.56262	4.20309
x_2^2	-0.0546935	0.01392	-3.9299	0.00025	-0.0826	-0.0268	-0.0826	-0.0268
x_3^2	-2.9301084	0.2326	-12.597	2E-17	-3.3968	-2.4634	-3.3968	-2.4634
x_4^2	-0.0074161	0.00116	-6.402	4.4E-08	-0.0097	-0.0051	-0.0097	-0.0051
x_5^2	-9.691E-06	1.3E-06	-7.4708	8.8E-10	-1E-05	-7E-06	-1E-05	-7E-06
x_6^2	-0.015199	0.00247	-6.1649	1.1E-07	-0.0201	-0.0103	-0.0201	-0.0103
x_1x_2	-0.0891158	0.09262	-0.9621	0.34044	-0.275	0.09675	-0.275	0.09675
x_1x_3	-0.6410141	0.53855	-1.1903	0.23935	-1.7217	0.43966	-1.7217	0.43966
x_2x_3	1.28191676	0.11731	10.9275	4.4E-15	1.04651	1.51732	1.04651	1.51732

P-value plays an essential role in assessing which input parameter is most significant for L-asparaginase production by SSF. Temperature, moisture, particle size, and cooking time are important input parameters in SSF as they significantly affect L-asparaginase activity as $p<0.05$. The quadratic effects of the operating conditions media, temperature, pH, moisture, particle size, and cooking time shows significant effect on L-asparaginase activity as p-value is reduced (<0.05). There is no significant influence on L-asparaginase activity ($p>0.05$) for the interaction between media & temperature, and media & pH, whereas temperature & pH interaction has a positive effect on L-asparaginase activity in SSF ($p<0.05$).

The scatter plot in fig. 1 (d) is depicting the relationship between experimental output and calculated output of enzyme activity obtained from the regression modeling. The value of the regression coefficient (R^2) between these two is 0.907. This value is quite higher, which suggest a strong linear relationship exist between the two and further validating the result predicted by the regression model.

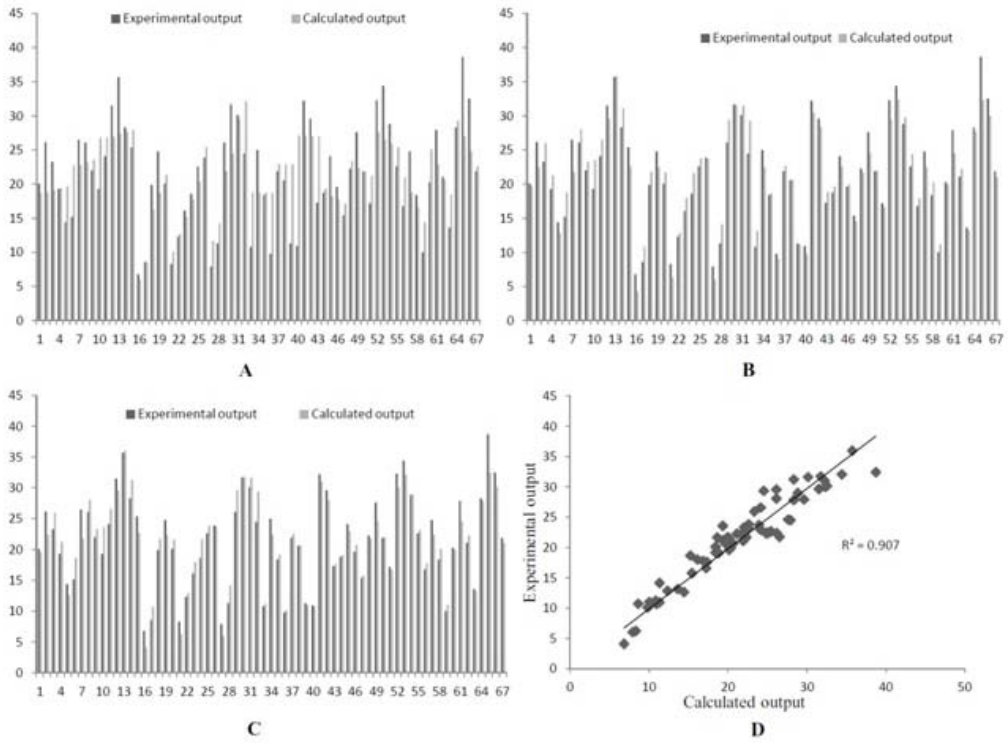

Fig. 1: (a) Comparison of experimental and calculated output after regression analysis when only linear terms were considered. (b) Comparison of experimental and calculated output after regression analysis when linear and quadratic terms were considered. (c) Comparison of experimental and calculated output after regression analysis when linear, quadratic and interaction terms were considered. (d) The scatter plot of experimental output and calculated output by the regression modeling to validate of data.

ANN results

In this study, the ANN simulations were performed for training, prediction and validation of L-asparaginase activity produced by SSF. There were six input variables (media, temperature, pH, moisture, particle size, and cooking time) that affect the enzyme activity and assigned as 6 corresponding input nodes. L-asparaginase activity was assigned as a single output node.

In fig. 2, neural network structures is shown with 6 input layers and two hidden layer for modeling operating conditions of L–asparaginase activity prediction.

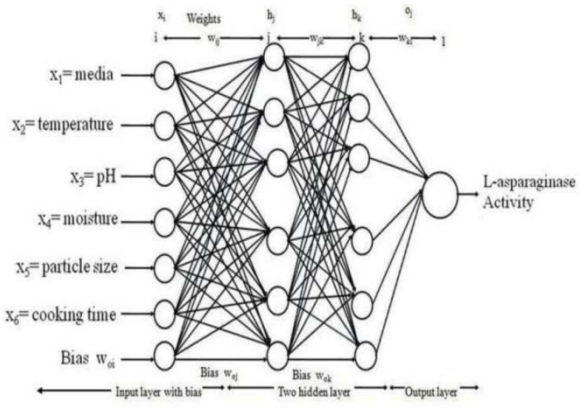

Fig. 2: Neural network structure with 6 input layers and two hidden layer for modeling operating conditions in prediction of L–asparaginase activity

There were total 67 experimental data used for ANN modeling. Among them, 57 data sets were employed to train the network and remaining 10 were utilized in testing and validation of the ANN model. While the training, network was performed SSE (sum of squared error) had been kept at 0.001 and the frequency of progress displays (in epochs) was set at 29 with maximum epochs of 1000. There may be different numbers of hidden layers, and the probable combination was tried to generate network architecture that provides the least error. The result of training, prediction and validation of L-asparaginase activity evaluation using the ANN model is shown in fig. 3. First, three plots of fig. 3 represent the performance of training, testing, and validation data and the fourth plot is of all the data used in ANN modeling. The dashed line in each plot represents the perfect result–outputs = targets and the solid line represents the best fit linear regression line between outputs and targets. The R^2 value indicates the relationship between the outputs and targets. Greater the value of R^2 (close to 1) more accurate is the linear relationship between outputs and targets. The value of R^2 is 0.996, 0.989 and 0.863 for training, validation and test data respectively which indicates a good fit. A similar study was carried by Baskar *et al.*, (2011) for *Enterobacter aerogenes* in submerged fermentation. Firstly, they have used Response Surface Methodology (RSM) and get R^2 value as 0.871 and when ANN back propagation algorithm was applied, the value of R^2 was 0.984. The predicted and experimental value of L-asparaginase activity was 18.59 and 18.72 IU/ml respectively in this study [12].

Fig. 4 represents the plot of performance by showing, training errors, testing errors, and validation errors. The plot shows that the final mean square error is small, indicating the result are reasonable. The training state is given in fig. 5.

The error histogram shows the data point where the fit significantly worse as compared to the majority of data (fig. 6). It is seen from the histogram that most errors fall between-1.729 &-1.499. There are test points with an error of-5.36 &-2.939 and validation with error of-3.343.

In predicting L-asparaginase activity, the back propagation neural network has proven very effective in a range of data taken in the learning set i.e. 57 data sets. With the help of simulation results, it was found that the introduction of two hidden layers has improved the forecasting performance of ANN as compared to the single hidden layer. The average value of MSE (Mean Square Error) was found to be reached 0.860, when only one hidden layer containing 10 neurons was used.

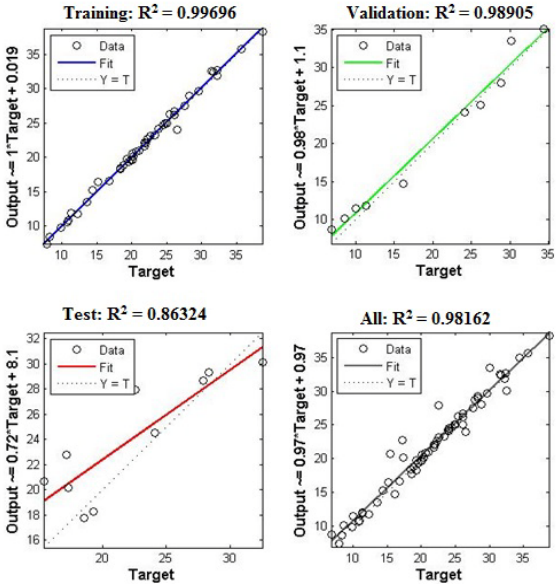

Fig. 3: Neural network Training, prediction and validation during ANN modeling for prediction of L–asparaginase activity in SSF

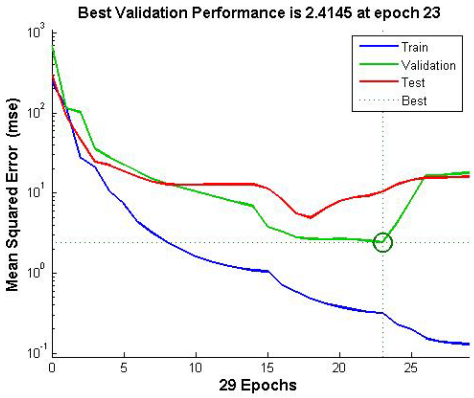

Fig. 4: Performance of training, testing and validation during ANN modeling for prediction of L–asparaginase activity in SSF

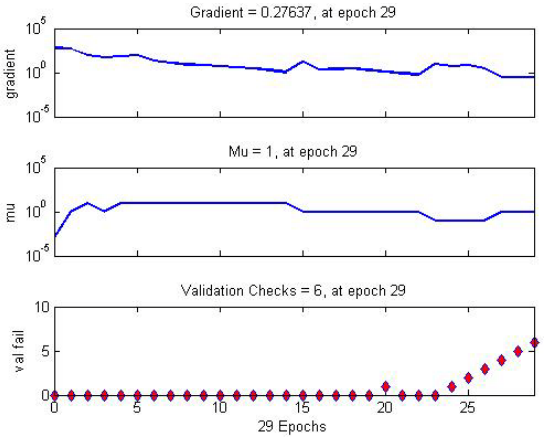

Fig. 5: Training state condition during ANN modeling for prediction of L–asparaginase activity in SSF

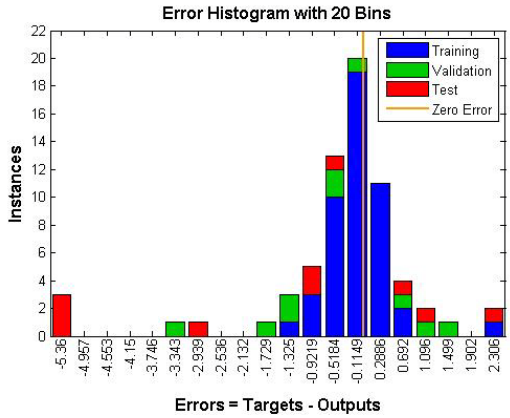

Fig. 6: Error histogram generated after ANN modeling representing the instances Vs errors

After that, it is decided to evaluate the network model using 10 testing data sets. The result obtained here says that the predicted L-asparaginase activity of ANN results are as close as with training and testing data provided to the network. In ANN modeling, there is a high value of the coefficient of determination (R^2 = 0.996) after training and with a low value of MSE as compared to full regression model (R^2 is 0.914) and standard error was 2.438. Thus, the ANN model is better fitted as compared to regression modeling. Gurunathan and Sahadevan in 2011 have performed another study, where evaluation and optimization of L-asparginase production by submerged fermentation using *A. terreus* was done [1]. They found the predicted and experimental value of enzyme activity was 36.64 and 36. 97 IU/ml respectively whereas, the value of R^2 determined was 0.995 [1]. They further expand their studies in the year 2012 with the same microorganism and through submerged fermentation but by changing the input parameters and get a higher value of enzyme activity as 40.86 IU/ml and 44.38 IU/ml by means of ANN [8].

The behavior of biological processes is quite difficult to understand and interpret. By applying ANN (a computational model), we have tried to implement the knowledge extracted from a biological system for accurate prediction of L-asparaginase enzyme activity. A huge number of data were generated at lab scale during the production process of an enzyme in SSF. In the present study, multi-linear regression and ANN approaches have been used to build a model for the cost-effective production of L-asparaginase. The reliable prediction was obtained by applying these approaches on different process parameters of L-asparaginase production, which will help in the enhancement of L-asparaginase productivity. These two statistical approaches were applied to the same problem for their comparative analysis. The R^2 represents the fraction of overall variance of the 'dependent' variable (L-asparaginase activity), which is explained by the 'independent' variable (input parameters: media, temperature, pH, moisture, particle size, and cooking time). The final value of R^2 during full regression model was 0.914, which represents 91.4% of the variation of the 'independent' variable can be explained by the variation of the 'dependent' variable. After doing the regression modeling, ANN tool was also applied further to investigate the relationship between input parameters and L-asparaginase activity. The back propagation network following the Levenberg-Marquardt technique was used for the training of network, and it efficiently predicts the L-asparaginase activity in SSF. After training, the trained network was tested using the experimental data to check further either the network has achieved good generalization or not. The accuracy of the ANN was assessed at each level from training to validating using the R^2 value and MSE value for every output. The high value of R^2 (0.996), R^2 (0.989) obtained in training & validation respectively, and low value means square error obtained in ANN modeling indicates that this model is efficient and accurate for significant prediction of L-asparaginase activity as compared to RM. The better performance obtained by ANN is because of its ability to capture non-linear dynamics. It is more beneficial in complex situations, where variables are constantly compared to convergence data, whereas the opposite is in the case of regression analysis. This study suggests the positive applicability of ANN for the prediction/optimization of enzyme activity and further attracts the interest of researchers to follow the same in future.

CONCLUSION

Due to high clinical and industrial importance of L-asparaginase, its huge demand is anticipated in future. L-asparaginase activity is significantly influenced by temperature, moisture, particle size, and cooking time in SSF. The effect of operating conditions on L-asparaginase activity produced by SSF was studied here using ANN and regression modeling because they are proved as a beneficial and accurate model for optimizing fermentation processes. This work would be precious for the judicious selection of different process parameter levels in SSF. It would significantly influence the improved production of L-asparaginase and efficient prediction of its activity for an imperative role in the economic production of this enzyme. The study demonstrated that hybrid ANN-RM modeling is an effective tool for understanding the influence of different critical factors of SSF. Similar modeling and optimization process may be used in future for the higher production and therapeutic application of biocatalyst, metabolites from suitable microorganisms.

ACKNOWLEDGEMENT

We are greatly thankful for the support provided by NIT Raipur (CG), India for providing the facility, space, and an opportunity for this work.

CONFLICT OF INTERESTS

Declared none

REFERENCES

1. Gurunathan B, Sahadevan R. Design of experiments and artificial neural network linked genetic algorithm for modeling and optimization of L-asparaginase production by *Aspergillus terreus* MTCC 1782. Biotechnol Bioprocess Eng 2011;16:50-8.
2. Kumar S, Dasu VV, Pakshirajan K. Purification and characterization of glutaminase-free L-asparaginase from *Pectobacterium carotovorum* MTCC 1428. Bioresour Technol 2011;102:2077-82.
3. Hymavathi M, Sathish T, Rao CS, Prakasham RS. Enhancement of L-asparaginase production by isolated *Bacillus circulans* (MTCC 8574) using response surface methodology. Appl Biochem Biotechnol 2009;159:191-8.
4. El-Bessoumy A, Sarhan M, Mansour J. Production, isolation, and purification of L-asparaginase from *Pseudomonas aeruginosa* 50071 using solid-state fermentation. J Biochem Mol Biol 2004;37:387-93.
5. Brown AM. A step-by-step guide to non-linear regression analysis of experimental data using a microsoft excel spreadsheet. Computer Methods Programs Biomedicine 2001;65:191-200.
6. Vats S, Negi S. Use of artificial neural network (ANN) for the development of bioprocess using *Pinus roxburghii* fallen foliages for the release of polyphenols and reducing sugars. Bioresour Technol 2013;140:392-8.
7. Rahman MBA, Chaibakhsh N, Basri M, Salleh AB, Rahman RNZRA. Application of artificial neural network for yield prediction of lipase-catalyzed synthesis of dioctyl adipate. Appl Biochem Biotechnol 2009;158:722-35.
8. Baskar G, Sahadevan R. Optimization of culture conditions and bench-scale production of L-Asparaginase by submerged fermentation of *Aspergillus terreus* MTCC 1782. J Microbiol Biotechnol 2012;22:923-9.
9. Mishra A, Production of L-Asparaginase an anticancer agent from *Aspergillus niger* using agricultural waste. Appl Biochem Biotechnol 2006;135:33–42.
10. Adamowski J, Chan HF, Prasher SO, Ozga-zielinski B, Sliusarieva A. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resour Res 2012;48:1–14.
11. Bolzan AC, Machado R, Piaia J. Egg hatchability prediction by multiple linear regression and artifical neural networks. Braz J Poult Sci 2008;10:97–102.
12. Baskar G, Rajasekar V, Renganathan S. Modeling and optimization of L-asparaginase production by *Enterobacter*

aerogenes using artificial neural network linked genetic algorithm. Int J Chem Eng Appl 2011;2:98–100.

13. Singh P, Singh S, Banik J, Mohan R. Optimization of cultural conditions using response surface methodology versus artificial neural network and modeling of L-glutaminase production by Bacillus cereus MTCC 1305. Bioresour Technol 2013;137:261–9.

How to site this article

- Archana Vimal, Jujjavarapu Satya Eswari, Awanish Kumar. Bioprocess modeling for the prediction of therapeutic enzyme L-asparaginase activity in solid state. Int J Pharm Pharm Sci 2016;8(7):420-426