Development of Process and Data Centric Inference System for Enhanced Production of L-Asparaginase From Halotolerant *Bacillus Licheniformis* PPD37

Payal Patel

Indrashil University

Haren B Gosai (✉ dh.haren@gmail.com)

Indrashil University

Haresh Panseriya

Gujarat Institute of Desert Ecology

Bharti Dave

Indrashil University

Research Article

Keywords: artificial neural network, process centric approach, halotolerant, L-asparaginase, response surface methodology

Posted Date: June 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-651645/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Applied Biochemistry and Biotechnology on November 30th, 2021. See the published version at <https://doi.org/10.1007/s12010-021-03707-5>.

Development of process and data centric inference system for enhanced production of L-asparaginase from halotolerant *Bacillus licheniformis* PPD37

Payal Patel¹, Haren Gosai^{1*}, Haresh Panseriya^{1,2}, Bharti Dave^{1*}

¹Department of Bioscience, School of Science, Indrashil University, Rajpur-Kadi, Dist. Mehsana, Gujarat, India-382740

²Gujarat Ecology Society, Synergy house, Subhanpura, Vadodara, Gujarat, India - 390003

Corresponding author email address: bpd8256@gmail.com; dh.haren@gmail.com

Abstract: The present study aims at bioengineering of medium components using – data and process centric approaches for enhanced production of L-asparaginase, an important biological molecule, by halotolerant *Bacillus licheniformis* PPD37 strain. To achieve this, first significant medium components were screened followed by optimization of a combination of media components and culture conditions such as L-asparagine, MgSO₄, NaCl, pH and temperature. Optimization study was carried out using statistical models such as response surface methodology (RSM) – process centric and artificial neural network (ANN) – data centric approaches. The production improved from 2.86 U/mL to 17.089 U/mL, an increase of approximately 6-times of the unoptimized L-asparaginase production. On comparing RSM and ANN models for optimised L-asparaginase production based on R² value, mean absolute percentage error (MAPE), root mean square error (RMSE), and mean absolute deviation (MAD) values, the ANN model emerged as the superior one. As this is the first report to the authors best knowledge on development of inference system using RSM and ANN models for enhanced L-asparaginase production using a halotolerant bacteria, this study could lead to more in-depth and large-scale L-asparaginase production.

Keywords: artificial neural network, process centric approach, halotolerant, L-asparaginase, response surface methodology

1. Introduction

L-asparaginase or L-asparagine amidohydrolase (EC 3.5.1.1) catalyses the deamidation of L-asparagine into L-aspartic acid and ammonia. This property of L-asparaginase has established its place as one of the most valuable biological macromolecules in the medical field and has been used for the treatment of acute lymphoblastic leukaemia (ALL) for nearly 50 years [1]. L-asparaginase functions as an enzymatic drug and depletes the L-asparagine present in the blood plasma, thus depriving the malignant tumour cells of it, ultimately causing their death by apoptosis [2]. Aside from ALL, L-asparaginase has found its use in the treatment of other types of cancer such as acute myeloblastic leukaemia (AML), non-Hodgkin's disease, melanoma, chronic lymphocytic leukaemia, Hodgkin's disease, reticulosarcoma, acute myelogenous leukaemia, and acute myelomonocytic leukaemia [3]. In the food industry, L-asparaginase is utilised to mitigate the formation of carcinogenic acrylamide during the processing of starchy food items at high temperatures due to Maillard's reaction [4].

The demand for L-asparaginase is expected to continue to rise and while commercial L-asparaginases (from *Escherichia coli* & *Erwinia chrysanthemi*) are widely available, they pose several questions that hinder their efficacy [5]. These asparaginases are known to induce a variety of side effects viz. hypersensitivity (anaphylaxis, pain, edema, rash, pruritus, erythema), pancreatitis, hyperglycaemia, thrombosis etc [6–10]. Thus, searching for a novel alternative source of L-asparaginase production with improved stability, less immunological response and high activity could resolve these problems. Halotolerant bacteria are yet to be thoroughly explored for enzyme production, particularly for their anti-cancer and anti-tumour activities. Microorganisms from hypersaline environment produce enzyme with different amino acid composition leading to a modified structure that enables them to withstand salinity and low amount of water [11]. L-asparaginase from such bacteria could possess unique properties that could make them suitable for treatment of hypersensitive patients and an attractive prospect for commercial production [12, 13].

Since fermentative production of any bioproduct is dependent on the media components and its cultural conditions such as pH, temperature, incubation time etc; the optimisation of the nutritional requirements and the operation conditions are of immense importance [14]. While the traditional method of optimising using One-Factor-At-Time (OFAT) is labour-intensive, time-consuming and prone to errors, response surface methodology (RSM) presents itself as a preferable option for inference system as it requires minimum number of experiments to predict best operating conditions [15]. RSM has the advantage of depicting accurate interaction between the multiple variables and how it affects the overall output i.e., the production [16]. Many studies have reported optimisation of bacterial enzyme production by using RSM method [17–19]. However, there is dearth of research in the optimisation of L-asparaginase production from halotolerant bacteria.

Apart from RSM, in recent years artificial neural network (ANN) has emerged as a popular tool for a wide range of bioprocess modelling and inference system including bioremediation and biodegradation techniques, quality assessment of water and sediments, forecasting of monsoon and river flow etc [20–23]. Although the use of ANN for optimisation of L-asparaginase production has been limited so far, no study on ANN analysis for halotolerant L-asparaginase has been reported to our best knowledge. Unlike traditional process centric approaches, ANN – a data centric approach gives good computational insights and require few parameters [24]. It was presumed that ANN need large number of datasets than RSM for modelling, but many reports suggest that ANN can work well even in limited number of datasets or input. Thus, it is possible to use RSM dataset to build even effective ANN fabricated model, using the same variables [25]. It is also assumed that ANN is able to enhance the accuracy between the non-linear input-output relationship. In contradictory to other computational and physics based numerical modelling, ANN requires training and calibration of the input variables. Thus, ANN infers the patterns on the basis of learning regime by use of input and output data without assuming or recognizing its nature and interrelations. Apart from that, there are many algorithms in ANN which offers flexibility for any additional limitations which may arise during training and calibration of the input [26]. Moreover, there is no specific design that is required, as decision making and drawing of conclusions is human intuitive [27]. Polynomial properties and regression based experimental modelling, makes ANN more suitable for designing of bioprocess model over other process

centric approaches [28]. ANN can predict all kind of non-linear functions including quadratic, while RSM is useful only for quadratic approximations [29].

Thus, the current study deals with development of inference system of process and data centric approaches i.e., RSM and ANN for optimisation of L-asparaginase production first, by designing and performing RSM and then using the experimental data obtained from RSM to build an effective ANN model with high accuracy. In addition to that, this study also emphasizes on their sensitivity analysis and their usefulness in optimisation of the process through the validation of the approaches. This experiment of model fabrication, prediction and successful application of traditional and non-linear models for L-asparaginase production by a halotolerant bacterium, is one of the first reports to the best of authors' knowledge.

2. Materials and Methods

2.1 Isolation and Screening of L-asparaginase producing halotolerant bacteria

Halotolerant *Bacillus licheniformis* PPD37 was isolated from the salterns of Sultanpura, Dandi situated at the coastline of the Arabian Sea, Gujarat, India and was selected for the present study. *B. licheniformis* PPD37 was cultured on HEM (Halophilic Enrichment Medium) modified from Kumar et al., 2012 at 37⁰C [31] and was found to have the highest L-asparaginase production out of the 51 total isolates (Data not shown) and hence used for subsequent optimization of medium components for enhanced production of L-asparaginase. Primary screening for L-asparaginase was done using rapid plate assay and the enzyme activity was derived by performing nesslerization method [32, 33].

2.2 Screening of MM9 medium components by Placket-Burman design (PB Design)

Screening of significant medium components of MM9 medium (inducing medium for L-asparaginase production) was performed by investigating components such as L-asparagine, Na₂HPO₄, NaCl, MgSO₄, KH₂PO₄, Glucose, and CaCl₂, based on Placket-Burman (PB) design. All the factors were examined at lower levels (-1) and higher levels (+1). The seven variables were screened in 13 experimental runs as observed in Table 1. 100 mL MM9 medium was prepared for the 13 experimental runs as designed by the algorithm in 250 mL Erlenmeyer flask and cultured at 37⁰C, 120 rpm for 72 hours. L-asparaginase production (U/mL) was estimated after every 24 hours by determining the amount of ammonia liberated through nesslerization method [33]. The experiments were

performed in duplicates and the average value of L-asparaginase activity was accepted as the response (Y). Media components with p value >0.05 were considered as significant. The statistical software Minitab v.19.0 was utilized for experimental design and analysis.

2.3 Experimental design and modelling

2.3.1 L-asparaginase production with screened parameters by RSM using CCD

RSM is a statistical method that optimises a process dependent on several variables, by predicting the ideal conditions that provide maximum yield with a minimum number of experiments. CCD were used in the present study is a type of RSM that fits a second-order polynomial order. The five variables selected to perform this experiment included three screened MM9 medium components: L-asparagine, NaCl and, MgSO_4 , as determined by PBD, and two physical conditions – pH and temperature. These factors were chosen to analyze their combined effect on the production of L-asparaginase [34–36]. All the input variables were examined at five levels and coded as -2, -1, 0, +1 and, +2 for lowest, low, medium, high and, highest concentration. As observed in Table 2, a total of 52 experimental runs (comprising of 32 cube points, 10 centre points in cube and, 10 axial points) were designed using Minitab v.19. Again, the experimental runs were carried out in duplicates in 250 mL Erlenmeyer flasks containing 100 mL of MM9 medium for 72 hours and the L-asparaginase activity (Y) was recorded after every 24 hours.

A second order quadratic model was developed to predict the influence of all variables on L- asparaginase production. The response surface quadratic equation for above 5 variables adopted in the present case was:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_{11}X_1^2 + b_{22}X_2^2 + b_{33}X_3^2 + b_{44}X_4^2 + b_{55}X_5^2 + b_{12}X_1X_2 + b_{23}X_2X_3 + b_{34}X_3X_4 + b_{14}X_1X_4 + b_{24}X_2X_4 + b_{13}X_1X_3 \quad \dots (1)$$

Where, Y is the desired response in terms of L-asparaginase production. b_0 is constant or intercept. $b_1, b_2, b_3, b_4, b_5, b_{11}, b_{22}, b_{33}, b_{44}, b_{55}$ and $b_{12}, b_{13}, b_{34}, b_{14}, b_{24}, b_{13}, b_{15}$ are the regression coefficients to be estimated. X_1, X_2, X_3, X_4 and X_5 are the coded values of the variables which were significant for each of the system.

Minitab v.19.0 was used for calculation of second-order polynomial coefficients, Analysis of Variance (ANOVA) and, regression analysis [37]. Fischer's Test at 95% confidence level was conducted to examine the significance of the model and goodness-of-fit for the second-order polynomial equation was determined by R^2 value. Contour

plots and 3-D graphs were used to visualize the interaction between significant factors.

2.4 Artificial neural network (ANN)

ANN is a type of statistical data modelling tool that can mimic different attributes of biological information processing and could be helpful in the optimisation process. For the development of an ANN model, the selection of a suitable learning method plays a crucial role as it reduces the error function and helps in improving the model [38]. The optimum conditions for L-asparaginase production were predicted by creating a linear feed-forward ANN model using MATLAB R2018a software [39–41]. The feed-forward model also known as multi-layer perceptron (MLP) has a gradient descent backpropagation (BP) learning algorithm and has been used expansively for biological applications [22, 40, 42]. BP algorithm performs this task by minimizing the error of changing weights that are inversely proportional to the negative error gradient.

2.4.1 Optimisation of network configuration

In an optimum, non-linear, multilayer perceptron (MLP) two layers exist – one input and one output with one or more layers hidden inside. Each layer has a precise number of artificial neurons and their estimation is dependent on the input/output vectors [22]. The determination of the hidden layer N_h is a crucial task for the optimisation of the ANN model and requires a trial-and-error based approach. Even though there is no particular formula for this purpose, the problem can be resolved by applying this thumb of rules: a) the number of the hidden layer (N_h) must be between I and $2I + 1$ and b) it should not be less than maximum of $I/3$ and O , where I = number of inputs and O = number of outputs [43]. Lesser neurons in the hidden layer are preferable as they are generally more efficient at generalization and have limited overfitting problems [44]. Since, in the present study, $I = 5$ and $O = 1$, the number of hidden neurons were found to be $\geq 5/3 = 1.66$ neurons and $\leq 5*4 = 20$ neurons implying that the hidden layer (N_h) lies between $1 \leq N_h \leq 20$. The trial-and-error approach was employed to finally determine the N_h based on the optimisation of neurons using MATLAB 2018a.

Except for the simulated data, the experimental data-set used for RSM-CCD analysis was again exercised by the ANN model to reproduce the L-asparaginase activity (U/mL) prediction (Table 2). In order to train the ANN model, the dataset was divided

into three subsets: 1) training, 2) validation and, 3) training, having 70%, 15% and, 15% of weightage, respectively. A predictive ANN model comprising of five parameters: L-asparaginase, MgSO_4 , NaCl, pH and, temperature as inputs and 20 neurons were constructed to obtain the output activity (U/mL) as illustrated by Fig 1. Since, we are taking a hybridised approach for optimisation, the experimental data obtained from RSM will be used as inputs to construct an ANN model. The dataset with different variables has multiple measurement units and are often weighed down by measurement errors and noise which may negatively influence the ANN training. By using activation function like logistic sigmoid, the raw dataset is standardized and converted into a non-dimensional form of uniform range of variability [23].

2.4.2 Determination of prediction potential of RSM and ANN

Several error functions and R^2 value were employed to ascertain the predictive ability of the RSM and ANN. During ANN training, the weights underwent appropriate adjustments to minimize the error functions. Most extensively utilized functions were applied for the study were mean squared error (MSE), root mean squared error (RMSE), mean absolute deviation (MAD) and mean absolute percentage error (MAPE). Microsoft Excel 2007 was used to calculate these errors by applying the following equations:

$$MAD = \frac{\sum_{i=1}^n |A_t - F_t|}{n} \dots\dots (2)$$

$$MSE = \frac{\sum_{i=1}^n (A_t - F_t)^2}{n} \dots\dots\dots (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (A_t - F_t)^2}{n}} \dots\dots (4)$$

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{A_t - F_t}{A_t} \right|}{n} \times 100 \dots (5)$$

Where, A_t refers to actual L-Asparaginase activity (U/mL), F_t is the predicted L-Asparaginase activity by RSM and ANN and n is the number of runs used for the experiment.

3. Results and Discussion

3.1 Screening of medium components by Plackett-Burman design (PBD)

PB design comprises of 13 experimental runs were used to observe the effect of the seven MM9 medium components. The components of MM9 media i.e. Na_2HPO_4 , KH_2PO_4 , L-asparagine, NaCl, MgSO_4 and, CaCl_2 were screened for production of L-asparaginase using *Bacillus licheniformis* PPD37. Table 1 represents the observed activity for every experimental run. The highest activity was observed from 3.62 to 10.84 U/mL after 48 hours of incubation which indicates clear impact of the medium components on L-asparaginase production. Vala et al., 2018 have observed maximum L-asparaginase production from fungi *Aspergillus niger* on 6th day of incubation after optimizing all the physico-chemical conditions. Thus, decrement in incubation time provides a very good opportunity for large scale production of this industrially important enzyme.

Three media components, L-asparagine, NaCl and, MgSO_4 were deemed to be significant factors based on their effect, p value (<0.05) and standard error (Table 3). By performing ANOVA with 95% confidence limit, p values of L-asparagine, NaCl and, MgSO_4 were calculated to be <0.0001 , 0.023 and 0.003, respectively. These medium components were used for the subsequent bioengineering for enhancement of L-asparaginase production through CCD-RSM.

3.2 CCD-RSM modelling

Optimisation of three screened medium components (L-asparagine, NaCl and, MgSO_4) and two physical conditions (pH and temperature) by RSM using CCD derived model predicted maximum L-asparaginase activity after 48 hours of incubation. This is in accordance with previous unoptimised run and Plackett-Burman screening, where the highest activity was observed after 48 hours and then plateaued or decreased after 72 hours. According to Table 2, the predicted values of L-asparaginase production closely matched to the experimentally calculated values.

The highest L-asparaginase activity of 17.089 U/mL (Table 2) was recorded after 48 hours. As revealed in Table 2, the maximum activity was obtained at L-asparagine – 2.0 g, MgSO₄ – 0.2 g, NaCl – 0.8 g, pH – 7.5 and temperature – 37 °C. The experimental and predicted values for activity seem to be near identical as depicted in Table 2. To verify the efficiency of the optimal run conditions, the process was repeated and the L-asparaginase activity was recorded to be 17.061 U/ml with specific activity 650 U/mg. *B. licheniformis* has been used for optimisation of L-asparaginase production by Mahajan et al., 2012; 2014, where they obtained enzyme activity of 32.26 U/ml however their specific activity was 23.1 U/mg which is significantly lower compared to the values obtained here [45, 46]. Alrumman et al., 2019 have also carried out L-asparaginase production using marine *B. licheniformis* and maximum enzyme production achieved was 8.1 U/mL for free cell culture and 11.66 U/mL for immobilized cells, with highest specific activity 36.08 U/mg [47]. Abdelrazek et al., 2019 have yielded 7.95 IU/mL of L-asparaginase production by RSM-Box-Behnken using *B. licheniformis*, which is comparatively lower [48].

Venil and Lakshmanaperumalsamy, 2008 have reported that addition of organic nitrogen sources in the media could induce the production of L-asparaginase [49]. Whereas, inorganic nitrogen source reduces the L-asparaginase production significantly. L-asparagine is the sole source of nitrogen in MM9 medium and it can work as an inducer for the production of L-asparaginase. Therefore, optimization of asparagine concentration may have significant effects on the enzyme production. This information was also supported by Sudhir et al., 2012 and Amena et al., 2010, where they have reported optimized production of L-asparaginase at 1% and 0.5% concentration of L-asparagine, respectively [50, 51]. A positive effect of pH on enzyme production was observed in the present study. Usually, at higher pH enzyme activity increases but this L-asparaginase was produced by marine bacteria, it may possible they are not susceptible at very high pH [52]. Therefore, higher production of L-asparaginase in the culture with pH 7.5 suggested positive impact of pH on L-asparaginase production after 48 hours.

ANOVA and regression analysis of the CCD is being represented in Table 4 & 5. On the basis of significant *p* value (< 0.05), L-asparagine, NaCl, temperature, L-asparagine*L-asparagine, NaCl*NaCl, pH*pH, temperature*temperature and, L-asparagine*NaCl were considered to be significant. Since the model has *p* value < 0.05 for

linear, square terms and 2-way interaction, it has been observed to be precise, significant, and reproducible. The actual R^2 value (0.9404) was found to be in a reasonable range from the adjusted R^2 (0.90) and predicted R^2 (0.80). The R^2 value of 0.94 is indicative of the 94.04% variability of the response explained by the model. After optimisation through CCD-RSM, L-asparaginase activity increased by approximately 6-times from 2.86 U/mL to 17.089 U/ml. The second-order response surface model equation fitted for L-asparaginase activity is as follows:

$$\begin{aligned} \text{Activity} = & -157.9 + 10.38 \text{ L - asparagine} - 48.4 \text{ MgSO}_4 + 4.60 \text{ NaCl} + 26.08 \text{ pH} \\ & + 4.246 \text{ temperature} - 1.209 \text{ L - asparagine} * \text{L - asparagine} \\ & + 58.9 \text{ MgSO}_4 * \text{MgSO}_4 - 0.749 \text{ NaCl} * \text{NaCl} - 1.987 \text{ pH} * \text{pH} \\ & - 0.06014 \text{ temperature} * \text{temperature} + 10.13 \text{ L - asparagine} * \text{MgSO}_4 \\ & - 1.497 \text{ L - asparagine} * \text{NaCl} - 0.196 \text{ L - asparagine} * \text{pH} - 0.0676 \text{ L} \\ & - \text{asparagine} * \text{temperature} - 6.73 \text{ MgSO}_4 * \text{NaCl} + 5.61 \text{ MgSO}_4 * \text{pH} \\ & - 0.420 \text{ MgSO}_4 * \text{temperature} + 0.117 \text{ NaCl} * \text{pH} - 0.0146 \text{ NaCl} \\ & * \text{temperature} + 0.0220 \text{ pH} * \text{temperature} \end{aligned} \quad \text{..... (6)}$$

Fig 2 represents the response surface of interactions between the five variables visualized through 3D graphs and contour plots. The graphs show the relationship between two variables and their effect on L-asparaginase activity after 48 hours of incubation. In the case of 3D graphs, the x and y axis represent the variables while z-axis corresponds to the response i.e., activity. Contour plots are the geometric illustrations of 3D graphs on a two-dimensional scale and the highest activity is depicted by the smallest curve or circle of the plot.

3.3 ANN modelling prediction

Input neuron network topology was utilised to construct an ANN architecture for the optimisation of L-asparaginase production. Out of numerous ANN topologies trained, the one that recognized the optimal number of neurons in the hidden layer, was selected based on the root mean square error (RMSE), mean absolute percentage error (MAPE) and, mean absolute deviation (MAD). The predictive ANN model constructed was trained by the BP algorithm since better transformation behaviour can be achieved through it. The BP algorithm uses second-order derivatives of the RMSE between the desired output and the actual output. For this study, the optimum ANN model topology (Fig. 1) has five input variables, one hidden layer with 20 neurons and, one output layer (5-9-1). As illustrated in Table 2, the predicted value of L-asparaginase activity for the ANN model is

similar to that of RSM. However, the higher value of R^2 (0.9846) along with lower value for RMSE (0.322) ascertains ANN as a preferable model for predicting L-asparaginase production conditions. This kind of results can be explained by the ability of ANN to provide global optimization, as training of the neurons was repeated several times for different physico-chemical variables. Repeated training could be helpful to provide global optimization of the variables [22]. A disadvantage of the ANN can be requirement of high-quality data for the training. In this context, if the experimental dataset is not normalized then variables with different measurement units can lead to various measurement errors and noise. These factors may exert negative impact on operation of ANN training algorithms.

3.4 Optimisation of the number of hidden neurons

The neurons present in the hidden layer plays a crucial role in accurate prediction of the optimal conditions. If the network topology is minimal i.e., having few neurons, the network is unable to train properly and learn. Thus, optimisation of the neurons was employed in order to determine the ideal neuron using the best predictivity and accuracy. The optimum neurons were selected on the basis of model performance including values of R^2 , RMSE, MAPE and MAD. Fig 1 illustrates the optimal structure of the feed forward network model for the neural network. As depicted in the Fig 3 the 11 neurons demonstrated the best prediction capability and high accuracy for L-asparaginase production. The results obtained gives a high value of R^2 (0.9846), indicating at reduction of RMSE (0.322) MAPE (1.519) and MAD (1.558) values, thereby, suggesting the significance of the ANN model. The obtained values further strengthen the capability of the model for predicting optimal neurons that yields maximum L-asparaginase production.

3.5 Training, validation and testing of the model

For the development of the model, the input data were compiled into three different subsets – training (70%), validation (15%) and, testing (15%). Here, Fig 4: a,b,c illustrates the ANN model having improved R^2 values in all three subsets, training (0.9824), validation (0.9953) and, testing (0.9876). Fig 4: d depicts that the overall model was a better fit to linear equation having R^2 value of 0.9846 higher than that of RSM (0.9404). The ANN model thus developed was able to accurately simulate L-asparaginase

production (target) and reproduce experimental output with superior precision. A BP algorithm trained, multi-layered feed forward ANN with substantial R^2 values can be utilized to obtain L-asparaginase production (target) with accurate results.

3.6 External validation of RSM and ANN models

Assessment of the prediction model is conducted on the basis of its competency to predict the outcome (response) with precision for new observations having distribution similar to that of training data. The performance of the constructed ANN prediction model was evaluated on the basis of explained variations (R^2) and on the probability of overfitting 'goodness-of-fit test. This probability may perhaps be eluded if the neurons are in a range of 8 to 12 [53]. Actual and predicted outputs of both RSM and ANN were contemplated for carrying out the external validation of the developed RSM and ANN models [54]. The results (Fig 5) showed that the ANN model displayed less overfitting in comparison with the RSM model, thereby, implying the ANN model as statistically more reliable and suitable.

3.7 Comparison of RSM and ANN model

3.7.1 Prediction Potential

Values of R^2 , RMSE, MAPE and, MAD for RSM and ANN models were calculated and compared to determine the model with superior predictive capability. Table 3 represents the experimental and predicted values achieved by RSM and ANN models. As seen in Table 6 the R^2 values for RSM and ANN are 0.9404 and 0.9846; RMSE values are 0.439 and 0.322; MAPE values are 2.632 and 1.519 and MAD values are 1.568 and 1.558, respectively. The ANN model established itself as superior over RSM, in every regard, for the production of L-asparaginase. The non-linear polynomials of the system exhibited by the ANN model could be accredited for its higher prediction potential; while RSM utilizes the quadratic equation to generalize data. Superiority of the ANN model over RSM as a modelling tool for bioprocess optimisation has been recorded several times, however, a limited number of reports have been found on the application of ANN for L-asparaginase production up till now.

3.7.2 Sensitivity Analysis

The study of medium components on L-asparaginase production is very important, especially interaction between medium components. This can be

meticulously studied using ANN compared to RSM. In RSM, the contribution of the various medium components can be measured by coefficients. Though, interpretation of ANN model can be difficult due to its 'black box' nature but by using 'Perturb' method it can give insights on the effectiveness of an individual parameter or interactions among parameters [55, 56]. Therefore, performance of the medium components was investigated by the optimal ANN model with 20 neurons in the hidden layer.

In RSM, the out of 5 variables, L-asparagine is the most influencing factor for L-asparaginase production. L-asparagine*MgSO₄, MgSO₄*pH and NaCl*pH had significant interactions between them indicating from their coefficient values (0.190, 0.105 and 0.0829) and suggesting their influencing role in the L-asparaginase production. On the other part, ANN can be useful to obtain insights between the medium components interaction due to the inherent nature of ANN. Fig. 7 indicates the rate of response and effect of this change in the output. The influence of the variable can be measured based on the slope and range of the change in response. As the value of slope increases, the influence of that individual variable also increases. It could be observed from the Fig 7 that the slope of the Asparagine (1.562) and temperature (0.716), suggesting greater influence on L-asparaginase production compared to other parameters viz. pH (-0.215), NaCl (-0.370) and MgSO₄ (-0.081). Interestingly, these results also supported by RSM first-order quadratic equation. Thus, it can be concluded that ANN is also equally efficient in sensitivity analysis.

In case of ANN, through intensive calibration, the low generalisation capability and erratic prediction can be improved to generate accurate predictions. Since ANN model is difficult to interpret due to its 'black box' nature, one need to analysis the noise generated through weights to gain a full understanding of the model. On the other hand, RSM could be useful for reduction in experimental runs or trials when compared to other traditional approaches like one factor at a time. Even though RSM is most commonly regarded method for optimization of bioprocesses, its use is limited to only quadratic non-linear correlation. Therefore, a hybrid model that combines both RSM and ANN could help in achieving successful optimisation.

4. Conclusion

In the present study, optimisation of physico-chemical conditions for enhanced L-asparaginase production by halotolerant *Bacillus licheniformis* PPD37 using two inference systems – RSM and ANN. RSM proved useful at determining the optimal conditions for L-asparaginase production. Highest activity of 17.089 U/mL, was observed at optimal condition of L-asparagine – 2.0 g, MgSO₄ – 0.2 g, NaCl – 0.8 g, pH – 7.5 and temperature – 37⁰C. A substantial rise of about 6-times from the previous one, was obtained by following the RSM model. Even though the predicted value for L-asparaginase activity given by both RSM and ANN is within a similar range (16.875 for RSM, 16.946 for ANN), ANN has proven to be better at prediction than RSM in this regard. ANN has an overall R² value of 0.9846 which is higher than 0.9404 of RSM, as well as lower values of other error functions. ANN has also proven to be equally useful for sensitivity analysis and has consistently performed better in every aspect. Therefore, ANN could potentially be a suitable alternative to RSM for the optimisation process. Thus, the study may lead to a more thorough exploration and research on the previously limited subject of L-asparaginase production from the halotolerant bacteria and improvement in large-scale production using such models. The study could be also helpful to design experiments for enhanced production of other important biological molecules which could be helpful to scientific community around the globe.

Acknowledgement:

Authors are thankful to Indrashil University for providing financial support and infrastructure for laboratory experiments to carry out this research. Authors are also thankful to Dr Imran Pancha, Assistant Professor, Department of Biology, SRM University – Amravati, Andhra Pradesh, India for providing suggestions for the improvement of the manuscript.

Funding

There is no external funding for the current research.

Conflict of Interest

Authors have no conflict of interest.

Author's Contribution Statement

Payal Patel – Methodology, formal analysis, writing, original draft,

Haren Gosai – Conceptualization, formal analysis, writing- review and editing

Haresh Panseriya - Writing- review and editing

Bharti Dave – Conceptualization, supervision, writing- review and editing

References:

1. Lanvers-Kaminsky, C. (2017). Asparaginase pharmacology : challenges still to be faced. *Cancer Chemotherapy and Pharmacology*, 79, 439–450. <https://doi.org/10.1007/s00280-016-3236-y>
2. Ueno, T., Ohtawa, K., Mitsui, K., Koder, Y., Hiroto, M., Matsushima, A., ... Nishimura, H. (1997). Cell cycle arrest and apoptosis of leukemia cells induced by L-asparaginase. *Leukemia*, 11, 1858–1861. <https://doi.org/10.1038/sj.leu.2400834>
3. Lopes, A. M., Oliveira-Nascimento, L. de, Ribeiro, A., Tairum, C. A., Breyer, C. A., Oliveira, M. A. de, ... Pessoa, A. (2017). Therapeutic l-asparaginase: upstream, downstream and beyond. *Critical Reviews in Biotechnology*, 37(1), 82–99. <https://doi.org/10.3109/07388551.2015.1120705>
4. Xu, F., Oruna-Concha, M. J., & Elmore, J. S. (2016). The use of asparaginase to reduce acrylamide levels in cooked food. *Food Chemistry*, 210, 163–171. <https://doi.org/10.1016/j.foodchem.2016.04.105>
5. Rangel-Yagui, C. de O., Costa-Silva, T. A., Santos, J. H. P. M., da Silva, F. V. S., Benyahia, B., Brumano, L. P., ... Junior, A. P. (2019). Development of L-Asparaginase Biobetters: Current Research Status and Review of the Desirable Quality Profiles. *Frontiers in Bioengineering and Biotechnology*, 6(January), 1–22. <https://doi.org/10.3389/fbioe.2018.00212>
6. Killander, D., Dohlwitz, A., Engstedt, L., Franzén, S., Gahrton, G., Gullbring, B., ... Åhström, L. (1976). Hypersensitive reactions and antibody formation during L-asparaginase treatment of children and adults with acute leukemia. *Cancer*, 37(1), 220–228. [https://doi.org/10.1002/1097-0142\(197601\)37:1<220::AID-CNCR2820370132>3.0.CO;2-W](https://doi.org/10.1002/1097-0142(197601)37:1<220::AID-CNCR2820370132>3.0.CO;2-W)
7. Pieters, R., Hunger, S. P., Boos, J., Rizzari, C., Silverman, L., Baruchel, A., ... Pui, C. H. (2011). L-asparaginase treatment in acute lymphoblastic leukemia. *Cancer*, 117(2), 238–249. <https://doi.org/10.1002/cncr.25489>

8. Kearney, S. L., Dahlberg, S. E., Levy, D. E., Voss, S. D., Sallan, S. E., & Silverman, L. B. (2009). Clinical Course and Outcome in Children With Acute Lymphoblastic Leukemia and Asparaginase-Associated Pancreatitis. *Pediatric Blood & Cancer*, 53, 162–167. <https://doi.org/10.1002/pbc.22076>
9. Iyer, R. S., Rao, S. R., Pai, S., Advani, S. H., & Magrath, I. T. (1993). L-asparaginase related hyperglycemia. *Indian Journal of cancer*, 30(2), 72—76. Retrieved from <http://europepmc.org/abstract/MED/8225380>
10. Beinart, G., & Damon, L. (2004). Thrombosis associated with L-asparaginase therapy and low fibrinogen levels in adult acute lymphoblastic leukemia. *American Journal of Hematology*, 77(4), 331–335. <https://doi.org/10.1002/ajh.20230>
11. Zolfaghar, M., Amoozegar, M. A., Khajeh, K., Babavalian, H., & Tebyanian, H. (2019). Isolation and screening of extracellular anticancer enzymes from halophilic and halotolerant bacteria from different saline environments in Iran. *Molecular Biology Reports*, (0123456789). <https://doi.org/10.1007/s11033-019-04787-7>
12. Ebrahiminezhad, A., Rasoul-Amini, S., & Ghasemi, Y. (2011). l-Asparaginase Production by Moderate Halophilic Bacteria Isolated from Maharloo Salt Lake. *Indian Journal of Microbiology*, 51(3), 307–311. <https://doi.org/10.1007/s12088-011-0158-6>
13. Shirazian, P., Asad, S., & Amoozegar, M. A. (2016). The potential of halophilic and halotolerant bacteria for the production of antineoplastic enzymes: L-Asparaginase and L-glutaminase. *EXCLI Journal*, 15, 268–279. <https://doi.org/10.17179/excli2016-146>
14. Baskar, G., Rajasekar, V., & Renganathan, S. (2011). Modeling and Optimization of L-asparaginase Production by Enterobacter Aerogenes Using Artificial Neural Network Linked Genetic Algorithm. *International Journal of Chemical Engineering and Applications*, 2(2), 98–100. <https://doi.org/10.7763/ijcea.2011.v2.83>
15. Sushma, C., Anand, A. P., & Veeranki, V. D. (2017). Enhanced production of glutaminase free L-asparaginase II by Bacillus subtilis WB800N through media optimization. *Korean Journal of Chemical Engineering*, 34(11), 2901–2915. <https://doi.org/10.1007/s11814-017-0211-1>
16. Kenari, S. L. D., Alemzadeh, I., & Maghsodi, V. (2011). Production of l-asparaginase from Escherichia coli ATCC 11303: Optimization by response surface methodology. *Food and Bioproducts Processing*, 89(4), 315–321. <https://doi.org/10.1016/j.fbp.2010.11.002>
17. Meena, B., Anburajan, L., Sathish, T., Vijaya Raghavan, R., Dharani, G., Valsalan Vinithkumar, N., ... Vijaya, R. (2015). L-Asparaginase from Streptomyces griseus NIOT-VKMA29: optimization of process variables using factorial designs and molecular characterization of L-asparaginase gen. *Scientific Reports*, 5(June), 131–139. <https://doi.org/10.1038/srep12404>
18. Dutta, J. R., Dutta, P. K., & Banerjee, R. (2004). Optimization of culture parameters for extracellular protease production from a newly isolated Pseudomonas sp. using response surface and artificial neural network models. *Process Biochemistry*, 39(12), 2193–2198. <https://doi.org/10.1016/j.procbio.2003.11.009>

19. Vasiee, A., Behbahani, B. A., Yazdi, F. T., & Moradi, S. (2016). Optimization of the production conditions of the lipase produced by *Bacillus cereus* from rice flour through Plackett-Burman Design (PBD) and response surface methodology (RSM). *Microbial Pathogenesis*, 101, 36–43. <https://doi.org/10.1016/j.micpath.2016.10.020>
20. Shamseldin, A. Y., Nasr, A. E., & O'Connor, K. M. (2002). Comparison of different forms of the Multi-layer Feed-Forward Neural Network method used for river flow forecasting. *Hydrol. Earth Syst. Sci.*, 6(4), 671–684. <https://doi.org/10.5194/hess-6-671-2002>
21. Khalil, B., Ouarda, T. B. M. J., & St-Hilaire, A. (2011). Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. *Journal of Hydrology*, 405(3), 277–287. <https://doi.org/https://doi.org/10.1016/j.jhydrol.2011.05.024>
22. Gosai, H. B., Sachaniya, B. K., Dudhagara, D. R., Panseriya, H. Z., & Dave, B. P. (2018). Bioengineering for multiple PAHs degradation using process centric and data centric approaches. *Chemometrics and Intelligent Laboratory Systems*, 179, 99–108. <https://doi.org/10.1016/j.chemolab.2018.04.019>
23. Dudhagara, D. R., Rajpara, R. K., Bhatt, J. K., Gosai, H. B., & Dave, B. P. (2016). Bioengineering for polycyclic aromatic hydrocarbon degradation by *Mycobacterium litorale*: Statistical and artificial neural network (ANN) approach. *Chemometrics and Intelligent Laboratory Systems*, 159(October), 155–163. <https://doi.org/10.1016/j.chemolab.2016.10.018>
24. Palani, S., Liong, S.-Y., & Tkalic, P. (2008). An ANN application for water quality forecasting. *Marine Pollution Bulletin*, 56(9), 1586–1597. <https://doi.org/https://doi.org/10.1016/j.marpolbul.2008.05.021>
25. Desai, K. M., Survase, S. A., Saudagar, P. S., Lele, S. S., & Singhal, R. S. (2008). Comparison of artificial neural network (ANN) and response surface methodology (RSM) in fermentation media optimization: Case study of fermentative production of scleroglucan. *Biochemical Engineering Journal*, 41(3), 266–273. <https://doi.org/https://doi.org/10.1016/j.bej.2008.05.009>
26. Çelekli, A., & Geyik, F. (2011). Artificial neural networks (ANN) approach for modeling of removal of Lanaset Red G on *Chara contraria*. *Bioresource Technology*, 102(10), 5634–5638. <https://doi.org/https://doi.org/10.1016/j.biortech.2011.02.052>
27. Ye, J., Zhang, P., Hoffmann, E., Zeng, G., Tang, Y., Dresely, J., & Liu, Y. (2014). Comparison of Response Surface Methodology and Artificial Neural Network in Optimization and Prediction of Acid Activation of Bauxsol for Phosphorus Adsorption. *Water, Air, & Soil Pollution*, 225(12), 2225. <https://doi.org/10.1007/s11270-014-2225-1>
28. Yasin, Y., Ahmad, F. B. H., Ghaffari-Moghaddam, M., & Khajeh, M. (2014). Application of a hybrid artificial neural network–genetic algorithm approach to optimize the lead ions removal from aqueous solutions using intercalated tartrate-Mg–Al layered double hydroxides. *Environmental Nanotechnology, Monitoring & Management*, 1–2, 2–7.

<https://doi.org/https://doi.org/10.1016/j.enmm.2014.03.001>

29. Buyukada, M., & Aydogmus, E. (2018). Utilization of apricot seed in (co-)combustion of lignite coal blends: Numeric optimization, empirical modeling and uncertainty estimation. *Fuel*, 216, 190–198. <https://doi.org/10.1016/j.fuel.2017.12.028>
30. Samuel, O. D., & Okwu, M. O. (2019). Comparison of Response Surface Methodology (RSM) and Artificial Neural Network (ANN) in modelling of waste coconut oil ethyl esters production. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*, 41(9), 1049–1061. <https://doi.org/10.1080/15567036.2018.1539138>
31. Kumar, S., Karan, R., Kapoor, S., Singh, S. P., & Khare, S. K. (2012). Screening and isolation of halophilic bacteria producing industrially important enzymes. *Brazilian Journal of Microbiology*, 43(4), 1595–1603. <https://doi.org/10.1590/S1517-83822012000400044>
32. Gulati, R., Saxena, R. K., & Gupta, R. (1997). A rapid plate assay for screening L-asparaginase producing micro-organisms. *Letters in Applied Microbiology*, 24(1), 23–26. <https://doi.org/10.1109/ASPDAC.2001.913268>
33. Imada, A., Igarasi, S., Nakahama, K., & Isono, M. (1973). Asparaginase and Glutaminase Activities of Micro-organisms. *Journal of General Microbiology*, 76(1), 85–99. <https://doi.org/10.1099/00221287-76-1-85>
34. Ashok, A., & Kumar, D. S. (2017). Different methodologies for sustainability of optimization techniques used in submerged and solid state fermentation. *3 Biotech*. Springer Berlin Heidelberg. <https://doi.org/10.1007/s13205-017-0934-z>
35. Darvishi, F., Faraji, N., & Shamsi, F. (2019). Production and structural modeling of a novel asparaginase in *Yarrowia lipolytica*. *International Journal of Biological Macromolecules*, 125, 955–961. <https://doi.org/10.1016/j.ijbiomac.2018.12.162>
36. El-Naggar, N. E. A., El-Ewasy, S. M., & El-Shweihy, N. M. (2014). Microbial L-asparaginase as a potential therapeutic agent for the treatment of acute lymphoblastic leukemia: The pros and cons. *International Journal of Pharmacology*. <https://doi.org/10.3923/ijp.2014.182.199>
37. Zafar, M., Kumar, S., & Kumar, S. (2010). Optimization of naphthalene biodegradation by a genetic algorithm based response surface methodology. *Brazilian Journal of Chemical Engineering*, 27(1), 89–99. <https://doi.org/10.1590/S0104-66322010000100008>
38. Stastny, J., & Skorpil, V. (2005). Neural networks learning methods comparison, 4, 325–330.
39. Gurunathan, B., & Sahadevan, R. (2012). Optimization of culture conditions and bench-scale production of L-asparaginase by submerged fermentation of *Aspergillus terreus* MTCC 1782. *Journal of Microbiology and Biotechnology*, 22(7), 923–929. <https://doi.org/10.4014/jmb.1112.12002>
40. Vala, A. K., Dudhagara, D. R., & Dave, B. P. (2018). Process-centric and data-centric strategies for enhanced production of l-asparaginase—an anticancer enzyme, using marine-derived *Aspergillus niger*. *Journal of Chemometrics*, 32(7), 1–11.

<https://doi.org/10.1002/cem.3024>

41. Turan, N. G., Mesci, B., & Ozgonenel, O. (2011). The use of artificial neural networks (ANN) for modeling of adsorption of Cu(II) from industrial leachate by pumice. *Chemical Engineering Journal*, 171(3), 1091–1097. <https://doi.org/https://doi.org/10.1016/j.cej.2011.05.005>
42. Pendashteh, A. R., Fakhru'l-Razi, A., Chaibakhsh, N., Abdullah, L. C., Madaeni, S. S., & Abidin, Z. Z. (2011). Modeling of membrane bioreactor treating hypersaline oily wastewater by artificial neural network. *Journal of Hazardous Materials*, 192(2), 568–575. <https://doi.org/https://doi.org/10.1016/j.jhazmat.2011.05.052>
43. Heaton, J. (2005). *Introduction to neural networks with Java*. Heaton Research Inc., Chesterfield.
44. Wang, L., Quek, H. C., Tee, K. H., Zhou, N., & Wan, C. (2005). Optimal size of a feedforward neural network: How much does it matter? In *Proceedings of the Joint International Conference on Autonomic and Autonomous Systems and International Conference on Networking and Services, ICAS/ICNS 2005* (p. 69). <https://doi.org/10.1109/ICAS-ICNS.2005.72>
45. Mahajan, R. V., Saran, S., Kameswaran, K., Kumar, V., & Saxena, R. K. (2012). Efficient production of l-asparaginase from *Bacillus licheniformis* with low-glutaminase activity: Optimization, scale up and acrylamide degradation studies. *Bioresource Technology*, 125, 11–16. <https://doi.org/10.1016/j.biortech.2012.08.086>
46. Mahajan, R. V., Kumar, V., Rajendran, V., Saran, S., Ghosh, P. C., & Saxena, R. K. (2014). Purification and characterization of a novel and robust L-asparaginase having low-glutaminase activity from *Bacillus licheniformis*: In vitro evaluation of anti-cancerous properties. *PLoS ONE*, 9(6). <https://doi.org/10.1371/journal.pone.0099037>
47. Alrumman, S. A., Mostafa, Y. S., Al-izran, K. A., Alfaifi, M. Y., Taha, T. H., & Elbehairi, S. E. (2019). Production and Anticancer Activity of an L-Asparaginase from *Bacillus licheniformis* Isolated from the Red Sea, Saudi Arabia. *Scientific Reports*, 9(1), 3756. <https://doi.org/10.1038/s41598-019-40512-x>
48. Abdelrazek, N. A., Elkhatab, W. F., Raafat, M. M., & Aboulwafa, M. M. (2019). Experimental and bioinformatics study for production of l-asparaginase from *Bacillus licheniformis*: a promising enzyme for medical application. *AMB Express*, 9(1), 39. <https://doi.org/10.1186/s13568-019-0751-3>
49. Venil, C. K., & Lakshmanaperumalasamy, P. (2012). Solid state fermentation for production of L – asparaginase in rice bran by *Serratia marcescens* SB08. *The Internet Journal of Microbiology*, 7(1). <https://doi.org/10.5580/1d06>
50. Sudhir, A. P., Agarwal, V. V., Dave, B. R., Patel, D. H., & Subramanian, R. B. (2016). Enhanced catalysis of l-asparaginase from *Bacillus licheniformis* by a rational redesign. *Enzyme and Microbial Technology*, 86, 1–6. <https://doi.org/10.1016/j.enzmictec.2015.11.010>
51. Amena, S., Vishalakshi, N., Prabhakar, M., Dayanand, A., & Lingappa, K. (2010). Production, purification and characterization of L-asparaginase from *Streptomyces gulbargensis*. *Brazilian Journal of Microbiology*, 41(1), 173–178.

<https://doi.org/10.1590/S1517-83822010000100025>

52. Yadav, S., Verma, S. K., Singh, J., & Kumar, A. (2014). Industrial Production and Clinical Application of L-Asparaginase: A Chemotherapeutic Agent. *International Journal of Biotechnology and Bioengineering*, 8(1), 54–60. <https://doi.org/10.5281/zenodo.1091554>
53. Choi, D.-J., & Park, H. (2001). A hybrid artificial neural network as a software sensor for optimal control of a wastewater treatment process. *Water Research*, 35(16), 3959–3967. [https://doi.org/https://doi.org/10.1016/S0043-1354\(01\)00134-8](https://doi.org/https://doi.org/10.1016/S0043-1354(01)00134-8)
54. Heidari, E., Sobati, M. A., & Movahedirad, S. (2016). Accurate prediction of nanofluid viscosity using a multilayer perceptron artificial neural network (MLP-ANN). *Chemometrics and Intelligent Laboratory Systems*, 155, 73–85. <https://doi.org/https://doi.org/10.1016/j.chemolab.2016.03.031>
55. Yetilmezsoy, K., & Demirel, S. (2008). Artificial neural network (ANN) approach for modeling of Pb(II) adsorption from aqueous solution by Antep pistachio (*Pistacia Vera* L.) shells. *Journal of Hazardous Materials*, 153(3), 1288–1300. <https://doi.org/10.1016/j.jhazmat.2007.09.092>
56. Jing, L., Chen, B., & Zhang, B. (2014). Modeling of UV-Induced Photodegradation of Naphthalene in Marine Oily Wastewater by Artificial Neural Networks. *Water, Air, & Soil Pollution*, 225(4), 1906. <https://doi.org/10.1007/s11270-014-1906-0>

Figures

Figure 1

Schematic representation of ANN model employed with input, hidden and output layers

Figure 2

Surface and contour plots visualizing the interactive effects of multiple variables on L- Asparaginase production by *B. licheniformis* PPD37 strain

Figure 3

Optimisation of neuron topology

Figure 4

Comparison between ANN-derived and experimental values of L-Asparaginase production

Figure 5

External validation of RSM and ANN predicted activity

Figure 6

Sensitivity analysis of ANN for variables

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [graphicalabstract.tiff](#)