



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

Modeling Lipase Production From Co-cultures of Lactic Acid Bacteria Using Neural Networks and Support Vector Machine with Genetic Algorithm Optimization

Sita Ramyasree Uppada¹, Aditya Balu², Amit Kumar Gupta², Jayati Ray Dutta^{1*}

¹ Biological Sciences Department, Birla Institute of Technology and Science, Pilani-Hyderabad Campus, Hyderabad, Andhra Pradesh, 500078, India.

² Department of Mechanical Engineering, Birla Institute of Technology and Science, Pilani-Hyderabad Campus, Hyderabad, Andhra Pradesh, 500078, India.

Abstract: Optimization of lipase production from co-cultures of *Lactococcus lactis* and *Lactobacillus plantarum*, *Lactococcus lactis* and *Lactobacillus brevis* was carried out. The lipase production was first modeled with Artificial Neural Network (ANN) and Support Vector Machine (SVM) taking various physico-chemical parameters (pH, temperature, incubation period, inoculum and substrate volume) into account. The yield obtained from SVM and ANN are compared based on correlation coefficient, % deviation from experimental results, computational time and the generalization capability. Later, the lipase production was optimized using Genetic Algorithm (GA). Experiments were performed on the process parameters, as obtained from GA and the results were validated to have % deviation to be less than 5%.

Keywords: Optimization; Microbial lipase; Co-culture; Artificial neural network; Genetic algorithm; Support vector machine.

I. Introduction

Lipases are versatile enzymes, which are used in food, dairy, detergent, pharmaceutical, leather, cosmetics, biosensor, pulp and paper industries[1]. They are important biocatalysts that perform various chemical transformations synthesizing a variety of stereo-specific esters, sugar esters, thioesters and amides[2]. Usually acid/base catalysts will catalyze organic chemical reactions, but due to their disadvantages like difficulty in recovery of byproduct, requirement of intensive energy for undergoing reactions and difficulty in removing catalyst, the reactions are accomplished by enzyme catalyst like lipases. The interest in microbial lipase production increased because of its huge industrial applications but the availability of genetically distinct lipases with specific characteristic is still limited and thus there is an immense need to develop stable lipases which can replace chemical catalysts. The primary importance in any fermentation process is an optimization of medium components[3]. Even a small increase in performance can have a significant impact on its production. Therefore, process optimization is one of the most frequently used operation in biotechnology. The classical method of optimization process was carried out by one factor at a time (OFAT) method by varying only a single factor and keeping the remaining factors constant[4]. This approach is not only time consuming, but also ignores the combined interactions between physico-chemical parameters[5]. Hence these classical techniques of optimization form a basis for developing advanced techniques more suitable to today's practical problems. Though microbial consortium of various microorganisms has been applied in many fields of biotechnology[6] its application for the production of lipase is yet to be explored more. It is well known that extracellular lipase production in microorganism is greatly influenced by physical factors like pH, temperature, incubation period, substrate volume and inoculum volume[7]. Therefore, considering the many industrial applications of lipase, we report here for the first time the optimization of extracellular lipase production using co-culture of *Lactococcus lactis* and *Lactobacillus plantarum*, *Lactococcus lactis* and *Lactobacillus brevis*.

II. Materials and Experimentation

A. Microorganism and lipolytic activity

For the present investigation, the co-cultures of *Lactococcus lactis* and *Lactobacillus plantarum*, *Lactococcus lactis* and *Lactobacillus brevis* were used for producing extracellular lipase. The strains were maintained on LB agar slants with a composition of tryptone (10g), yeast extract (5g), NaCl (10g), agar (15 g) and distilled water (1lt) at 4°C. For preliminary screening of lipase producing bacteria, tributyrin agar was used. All the cultures were inoculated into tributyrin agar plates containing peptone (5g), beef extract (3g), tributyrin (10ml), agar-agar (20g) and distilled water (1lt) and kept for incubation at 37°C for 24 hours and observed for zone formation. A clear zone around the colonies indicated the production of lipase.

B. Enzyme assay

The lipase assay was performed spectrophotometrically using p-nitro phenyl palmitate as substrate. The assay mixture contained 2.5 ml of 420µm P-nitrophenol palmitate, 2.5ml of 0.1 M Tris – Hcl (pH-8.2) and 1ml of enzyme solution. It was incubated in water bath at 37°C for 10 min. p-nitrophenol was liberated from p-

nitrophenylpalmitate by lipase mediated hydrolysis imparting a yellow color to the reaction mixture. After incubation, the absorbance was measured at 410 nm [8]. One unit (U) of lipase was defined as the amount of enzyme that liberates one micromole of p-nitro phenol per minute under the assay conditions.

C. Experimental design and Lipase production

The optimum levels for extracellular lipase production by the *L. lactis* and *L. brevis*, *L. lactis* and *L. plantarum* strains with respect to incubation period, temperature, pH inoculum volume and substrate volume, were obtained by single factor optimization by conducting the experiments in 250 ml Erlenmeyer flasks containing 50 ml of medium comprising of peptone (0.5%), yeast extract (0.3%), NaCl (0.25%), MgSO₄ (0.05%) with olive oil as a substrate and inoculated with the freshly prepared bacterial suspension at 35°C. After incubation, the cell-free supernatant was obtained by centrifugation at 7197xg for 20 minutes and the extracellular lipase activity of the fermented broth was determined. Experiments were conducted in triplicate and the results were the average of these three independent trials. Table 1 shows the chosen process parameters and their levels used.

Table 1. Process parameters and levels of the experiment

Level	pH	Temperature (°C)	Incubation Period (hrs)	Inoculum volume(ml)	Substrate volume(ml)
1	5	25	24	0.5	0.5
2	5.5	30	36	1	1
3	6	35	48	1.5	1.5
4	6.5	40	60	2	2
5	-	-	72	-	-
6	-	-	84	-	-

In the next stage, ANN and SVM models are built to study the interactive effects of the five variables, i.e. pH, temperature, inoculum volume, incubation period, substrate volume.

III. Artificial Neural Network Model

Machine learning provides tools that automate the computer to recognize the complex patterns and make intelligent decisions based on data. One of the very useful tools in the field of the Machine Learning is the Artificial Neural Networks [9]. Predictive modeling of physical law which is highly complex is often done using the tool of ANN [10]. The most basic unit of ANN is a neuron. A neuron has a similar function like a biological neuron, it combines all the inputs given to the neuron and transfers it to another neuron based on an activation function. Tan sigmoid, linear line and Log sigmoid are the popularly used activation functions, also known as transfer functions. A tan sigmoid function is defined as follows:

$$f(t) = \frac{e^t - e^{-t}}{e^t + e^{-t}} \quad (1)$$

Where, t represents the input to the tansigmoid function. Similarly, log sigmoid and linear line are also defined. A group of neurons connecting together in a weighted form gives rise to output is called a layer of neurons. In general, there are many layers in a neural network. The layer taking the actual input is called the input layer and the layer which ultimately gives the output is called the output layer. The intermediate layers are called hidden layers. The collection of layers is known as a neural network. The number of neurons in the input layer and output layer is fixed to the model which is built. However, the number of neurons in the hidden layer is a variable. Evaluation of the weights of every layer in the neural network is known as training of the neural network. The back propagation training algorithm is most popularly used for training neural network and to estimate the values of the weights. Hence, in this paper, back propagation algorithm is used for training the feed forward neural networks. The Differential Evolution algorithm (DE) was used for training the networks and for the tuning of weights so as to determine the optimal architecture of neural networks. The neural network toolbox of the MATLAB software package is used for training, testing and validation of the given data.

IV. Support Vector Machine Model

Support Vector Machines (SVMs) is most important method of supervised learning, which analyses and recognizes data patterns, useful for classification and regression [11]. The advantage with this type of algorithm is easy attainment of the global minimum, and avoiding of the local minimum as in other methods such as neural networks [12]. Algorithm performance significantly depends on the choice of kernel function which maps the input space to the transformed feature space. Several non-linear-mapping functions exist to do such conversion. These functions are known as kernel function. One of the most popularly used kernel functions is Radial Basis Function (RBF).

N training data $\{(x_i, y_i)\}_{i=1}^N \in R^m \times R$ with x_i as the input vector and y_i as the actual output value is used for the SVM model. SVM model is expressed as follows:

$$y = f(x) = \sum_{i=1}^N w_i \phi_i(x) + b = w^T \phi(x) + b \quad (2)$$

Where the function $\phi_i(x)$ is called the non-linearly mapped feature from the input space x , $w = [w_1 w_2 \dots w_N]^T$ is the weight vector and $f = [f_1 f_2 \dots f_N]^T$ is the basis function vector. The result of the model is a hyper-surface which is a non-linear surface. However, it is converted into a linear regression model by mapping the

input vectors x to vectors $f_i(x)$ of a high dimensional kernel-induced feature space. The parameters w and b are support vector weight and a bias that is calculated. The learning task is defined as the minimization problem of the regularized risk function (L). The kernel function chosen learns and minimizes the regularized risk function.

$$L(p) = \frac{1}{2} p^T p + \lambda \sum_{i=1}^N |y_i - f(x)|_e \quad (3)$$

$$\text{where } |y_i - f(x)|_e = \begin{cases} 0, & \text{if } |y_i - f(x)| < \varepsilon \\ |y_i - f(x)| - \varepsilon, & \text{otherwise} \end{cases} \quad (4)$$

The two variables ε and λ are the parameters that control the dimension of the approximating function. Both must be selected by the user. By increasing the insensitivity zone, governed by ε , the accuracy of the approximation is reduced, which decreases the number of support vectors, leading to data compression. In addition, while modelling highly noisy and polluted data, increasing the insensitivity zone ε has smoothing effect. The regularization parameter (λ) determines the tradeoffs between the weight vector norm and the approximation error. Hence the parameters need to be fine-tuned using any optimization techniques. One such technique is DE algorithm. LS-SVM (least square-support vector machines) toolbox built in MATLAB environment is applied to the data obtained from the experiments. The parameters λ and ε are determined using the DE algorithm. Firstly, the DE starts with the initial values of $\lambda = 100000$ and $\varepsilon = 0.1$ and then the evolutionary DE algorithm which uses cross validation to fine tunes the parameters. The kernel function used is the radial basis function which is used extensively in the literature.

V. Optimization of Lipase production using Genetic Algorithm

A Genetic algorithm is a stochastic optimization technique that searches for an optimal value of a complex objective function and are used to solve complicated optimization problems by simulation or mimicking a natural evolution process[13]. GA has been successfully used as a tool in computer programming, artificial intelligence, optimization[14] and neural network training and information technology. In a GA, a population of candidate solutions (called chromosomes) to an optimization problem is evolved towards fitter solutions in an iterative process. Each candidate solution is mutated and altered; traditionally, solutions are represented in binary as strings of 0s and 1s, but other encodings are also possible. The selection of chromosomes for the next generation is called reproduction, which is determined by the fitness of an individual. Different selection procedures are used in GA depending on the fitness values, of which proportional selection, tournament selection and ranking are the most popular procedures[15]. In this study, the settings for GA in MATLAB are as follows which is explained in Table 2.

Table 2. The parameters used for optimization using GA

GA Algorithm	
Population size	100
Generations	100
Crossover Probability	0.8
Mutation function	Constraint dependent
Elite count	2
Max. Function Evaluations	100000

VI. Results and Discussion

The ANN and SVM models for the lipase production were trained as explained in the algorithm above. The correlation coefficient is one of the statistical measures used to judge the goodness of fit of a model. The correlation coefficient between two variables X and Y is measured using the Pearson product-moment coefficient, which takes the value between -1 and +1 inclusive. It is defined by the formula:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

The X_i represents the original values as obtained in the experiments. $\bar{X}$ is the mean of the original values. Similarly, Y represents the predicted values. The ideal prediction is supposed to give a value of r which is equal to one. Consequently, the ideal prediction leads to a straight line with slope 1, as the X-axis and Y-axis represent the experimental and predicted values by each of the methods employed. The complete data are divided into training, testing and validation data (80%, 10% and 10% respectively). The training data are used for the training of the ANN and SVM models. The validation data is used to determine the optimal or the best ANN architecture or best SVM parameters for good fit with the experimental results. The testing data is used to finally compare the ANN and SVM models based on the their predictive capability at unknown process parameters. The correlation plots of training, testing and validating data of ANN and SVM for co-culture of *Lactococcus lactis* and *Lactobacillus plantarum* are shown in Figs. 1 (a-c) and 2 (a-c) respectively.

Fig.1 The plot of Predicted vs. Experimental (a) training data (b) testing data (c) validation data of ANN model for lipase production from co-cultures of *Lactococcus lactis* and *Lactobacillus plantarum*

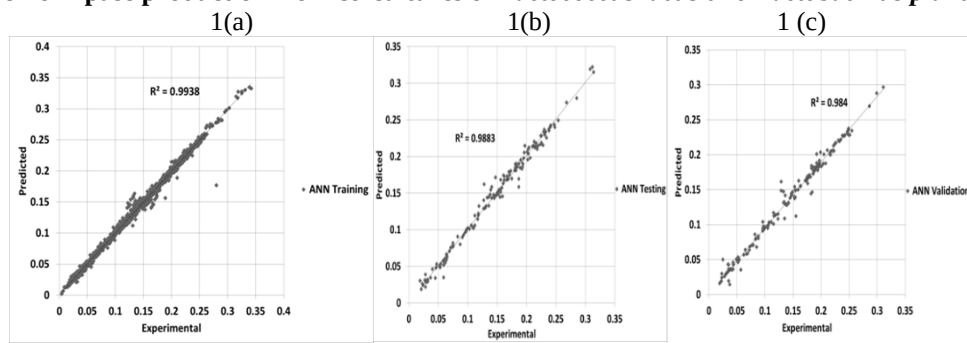
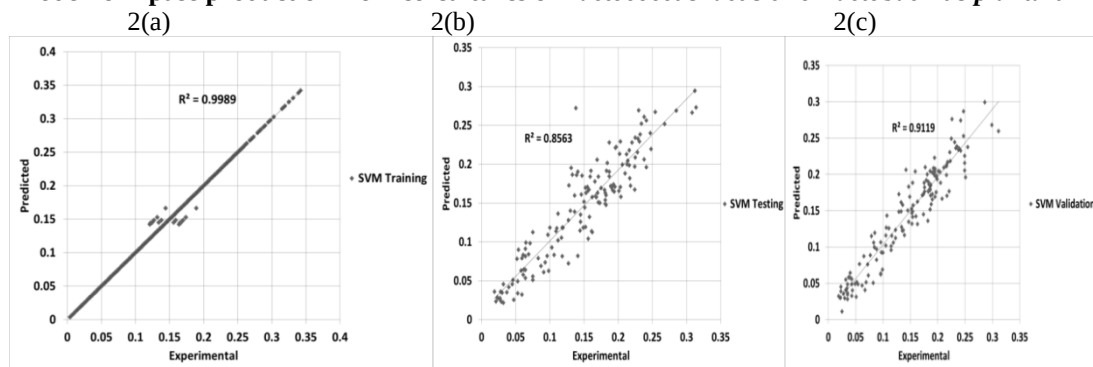


Fig. 2 The plot of Predicted vs. Experimental (a) training data (b) testing data (c) validation data of SVM model for lipase production from co-cultures of *Lactococcus lactis* and *Lactobacillus plantarum*



Similarly, the correlation plots of training, testing and validating data of ANN and SVM for co-culture of *Lactococcus lactis* and *Lactobacillus brevis* are shown in Fig. 3 (a-c) and 4 (a-c) respectively.

Fig. 3 The plot of Predicted vs. Experimental (a) training data (b) testing data (c) validation data of ANN model for lipase production from co-cultures of *Lactococcus lactis* and *Lactobacillus brevis*

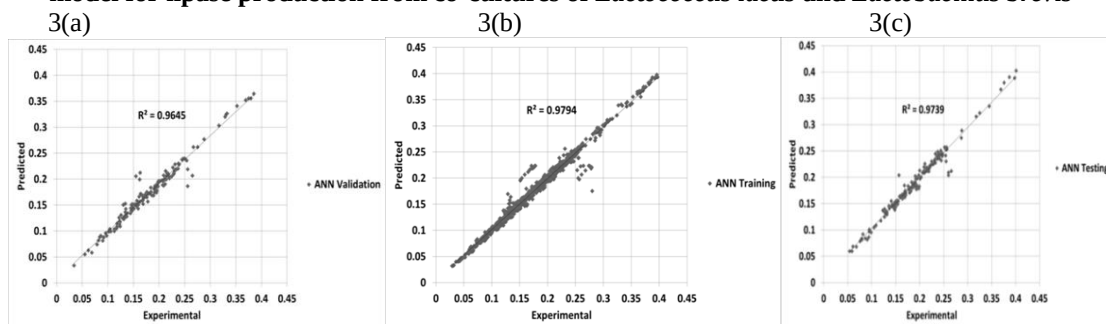
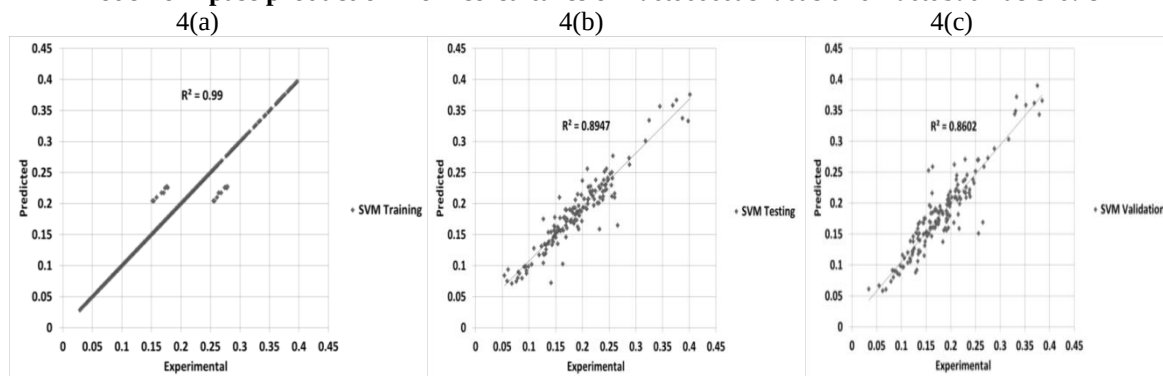


Fig. 4 The plot of Predicted vs. Experimental (a) training data (b) testing data (c) validation data of SVM model for lipase production from co-cultures of *Lactococcus lactis* and *Lactobacillus brevis*



It is seen in general from these figures that SVM has good correlation with the experimental results for the training data set. Hence, it is seen that ANN is having more predictability of yield from lipase production due to its

good correlation with validation and testing data. Later, the lipase production is optimized using the GA toolbox as mentioned in the previous sections. The optimal yield as obtained from the GA from ANN and SVM models of lipase production are shown in Tables 3 and 4

Table 3. Settings of parameters and predicted yield obtained for co-culture of *Lactococcus lactis* and *Lactobacillus plantarum*

Co-culture of <i>Lactococcus lactis</i> and <i>Lactobacillus plantarum</i>								
Model	pH	Temp(^o C)	Input parameters			Experimental Yield	Predicted Yield	% deviation
			Incubation (hrs)	Inoculum volume (ml)	Substrate volume (ml)			
SVM	5.492	34.828	76.230	1.768	1.760	0.414	0.428	3.365
ANN	5.450	34.944	77.213	0.518	2.000	0.400	0.412	2.887

Table 4. Settings of parameters and predicted yield obtained for co-culture of *Lactococcus lactis* and *Lactobacillus brevis*

Co-culture of <i>Lactococcus lactis</i> and <i>Lactobacillus brevis</i>								
Model	pH	Temp.(^o C)	Input parameters			Experimental Yield	Predicted Yield	% deviation
			Incubation Period (hrs)	Inoculum volume (ml)	Substrate volume (ml)			
SVM	5.658	25.019	72.980	1.486	1.997	0.350	0.361	3.288
ANN	5.574	29.934	60.295	1.695	1.799	0.362	0.369	1.869

The input parameters represent the process parameters using which the optimal yield was obtained. The predicted yield is the yield at optimum settings, obtained from the GA toolbox for the ANN and SVM models. The Experimental yield represents the experimental results obtained. The final column represents the deviation in the yield from the experimental value found out by

$$\% \text{ deviation} = \frac{|\text{experimental yield} - \text{predicted yield}|}{\text{predicted yield}} \times 100$$

The % deviation in the value is less than 5% in all the models obtained.

Further, it is observed that the % deviation from ANN model is lesser than the SVM model. This suggests that ANN is a better predictive model due to its lesser % deviation, better correlation with the experimental results. However, there are a few more considerations which have been identified in the literature. The computational time taken from ANN is 122.93 seconds, whereas that of SVM is 0.45 seconds. Thus, it was seen that SVM takes less computational time than ANN. Similar results were observed with the lipase production results. The mean computational time taken for ANN was 1.63 seconds and 24.1 seconds for ANN and SVM respectively. Further, it seen that the prediction of lipase production is significantly accurate. The % deviation from the predicted results is under 5%. Hence, it can be concluded that both ANN and SVM are having comparable results. The higher yield obtained from SVM model is justified due to the generalization capability. When a model is allowed to completely predict the results based on the present experimental results, there is a general tendency to incur more error at unknown data points. However, if the data is allowed have little bit of error in the prediction with the present experimental data points, the tendency to overfit the data is less(this phenomenon is known as generalization capability as explained by Cristopher and Bishop). Hence, it can be seen that SVM is better than ANN is terms of computational time and generalization capability. Whereas ANN is better than SVM in terms of correlation coefficient and % deviation.

VII. Conclusion

In this research study, ANN and SVM models were built using fermentation performance parameters (pH, temperature, incubation period, inoculum and substrate volume) to predict the extracellular lipase production from co-cultures of *Lactococcus lactis* and *Lactobacillus plantarum*, *Lactococcus lactis* and *Lactobacillus brevis*. The following two conclusions can be drawn from the results.

1. Considering the computational time and the generalization capability of both the models, it was found that SVM was better than the ANN model
2. Considering the % deviation and the correlation coefficient with the experimental data, it can be found that ANN is better than SVM.

Further, it is also seen that on application of GA, the lipase production yield is optimized and the results are validated with experiments at the input parameters obtained from GA for optimal yield.

References

- [1] M.P. Licia, M.R. Mario and R.C. Guillermo, "Catalytic properties of lipase extracts from *Aspergillus niger*," Food Technol. Biotech, vol. 44, March 2006, pp. 247-252.

- [2] B. S. Sooch and B. S. Kauldhar, "Influence of multiple bioprocess parameters on production of lipase from *Pseudomonas* sp. BWS-5," *Braz. Arch. Biol. Technol.*, vol.56, Sept/Oct. 2013, pp. 711-721.
- [3] Y.R. Abdel- Fattah, N.A.Soliman, S.M. Yousef and E.R. El-Helow, "Application of experimental designs to optimize medium composition for production of thermostable lipase/esterase by *Geobacillus thermodenitrificans* AZ1, *J. Genet Eng. Biotechnol.*, vol. 10, Dec. 2012, pp. 193–200.
- [4] S. Ghosh, S. Murthy, S. Govindasamy and M. Chandrasekaran, "Optimization of L-asparaginase production by *Serratia marcescens* (NCIM 2919) under solid state fermentation using coconut oil cake, *Sustain. Chem. Process*, vol.1, March 2013, pp.1-8.
- [5] P. Kanmani, S. Karthik, J. Aravind and K. Kumaresan, "The use of Response surface methodology as a statistical tool for media optimization in lipase production from the Dairy effluent isolate *Fusarium solani*," *ISRN Biotechnology*, vol. 2013, Sep 2012, pp. 8 .
- [6] R. Kaushal, N. Sharma and D. Tandon, "Cellulase and xylanase production by co-culture of *Aspergillus niger* and *Fusarium oxysporum* utilizing forest waste, *Turk. J. Biochem*, vol. 37, March 2012, pp.35–41.
- [7] A. Saha and S. C. Santra, "Isolation and Characterization of Bacteria Isolated from Municipal Solid Waste for Production of Industrial Enzymes and Waste Degradation, *J. Microbiol. Exp*, vol.1, May 2014, pp.1-8.
- [8] N.Verma, S.Thakur and A.K.Bhatt, "Microbial Lipases: Industrial Applications and Properties (A Review)," *Int. Res. J. Biological Sci*, vol.1, Dec 2012, pp.88-92.
- [9] C.H. Kuo, T.A. Liu, J.H. Chen, C.M. Chang and C.J. Chieh, "Response surface methodology and artificial neural network optimized synthesis of enzymatic 2-phenylethyl acetate in a solvent-free system," *Biocatal. Agric Biotechnol*, vol. 3, July 2014, pp.1-6.
- [10] B. Fathiha, B.Sameh, S. Yousef, D. Zeineddine and R. Nacer, "Comparison of artificial neural network (ANN) and response surface methodology (RSM) in optimization of the immobilization conditions for lipase from *Candida rugosa* on Amberjet®) 4200-Cl, *Prep Biochem Biotechnol*, vol. 43, Dec 2012, pp. 33-47.
- [11] G. Zhang and H. Ge, "Prediction of xylanase optimal temperature by support vector regression," *Electron. J. Biotechnol*, vol. 15, January 2012, pp.8.
- [12] L. Morgado, C. Pereira, P. Verissimo and A.Dourado, "Modelling proteolytic wnzymes with Support vector machines," *J. Integrative Bioinformatics*, vol. 8, Dec 2011, pp.170.
- [13] M. Chauhan, R.S. Chauhan, and V.K. Garlapati, "Modelling and optimization Studies on a novel lipase production by *Staphylococcus arlettae* through submerged fermentation," *Enzyme Research*, vol. 2013, Nov 2013, pp. 8.
- [14] A. Sheta, R. Hiary, H. Faris and N. Ghatasheh, "Optimizing thermostable enzymes production using Multigene symbolic regression genetic programming," *World Applied Sciences Journal*, vol. 22, April 2013, pp.485-493.
- [15] S. R. Uppada, A. k. Gupta, and J.R. Duta, "Statistical optimization of culture parameters for lipase production from *Lactococcus lactis* and its application in detergent industry, *Int. J. ChemTech Research*, vol. 4, Oct-Dec 2012, pp. 1509-1517.