World Applied Sciences Journal 22 (4): 485-493, 2013
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2013.22.04.7694

Optimizing Thermostable Enzymes Production Using Multigene Symbolic Regression Genetic Programming

[1]Alaa Sheta, [2]Rania Hiary, [3]Hossam Faris and [4]Nazeeh Ghatasheh

[1]Computer Science Department, College of Computers
and Information Technology, Taif University, Taif, Saudi Arabia
[2]Information and Communication Technology in Education Program,
Al-Albait University, Mafrag, Jordan
[3]King Abdulla II School for Information Technology,
The University of Jordan, Amman, Jordan
[4]Business Information Technology Department,
The University of Jordan, Aqaba, Jordan

Abstract: Thermostable enzymes production depends on number of attributes such as temperature, pH, inoculum, time and agitation. Optimizing the relationship between these attributes has been a challenge in biochemical research field. Machine learning techniques such as Artificial Neural Networks (ANN), Fuzzy Logic (FL) and Genetic Algorithms (GAs) were used to solve the lipase activity modeling problem. In this paper, we explore the use of Multigene Symbolic Regression GeneticProgramming to solve the production problem of a solvent, detergent, and thermotolerantlipase using the Newly IsolatedAcinetobacter sp. in submerged and solid-state fermentation. Five attributes will be used to develop a mathematical model for the lipase activities. They are temperature, pH, inoculum, time and agitation. Genetic Programming shows promising results compared to reported results in the literature.

Key words: Multigene Genetic Programming · Symbolic regression · Optimization · Thermostable Enzymes

INTRODUCTION

Recently, there has been a growing interest in investigating the features of thermostable enzymes not only because of their extra thermostability but often due to their extra resistance to many environments changes than their mesophilic homologues [1-4]. Diversity of thermostable (TS) enzymes has been successfully used in industrial applications, mainly as replacements for thermolabile (TL) enzymes [5-8]. On the industrial production scale, microbial extracellular enzymes show remarkableadvantages especially in biotechnological applications such as diary-based products, detergents, drugs, cosmetics and leather processes [9-11]. Many operating conditions which include nutritional and physico-chemical factors such as temperature, initial pH, incubation period, time, inoculum size and agitation rate,highly affect the production of thermostable lipase enzymes [4, 11].

Recently, lipase production is getting more and more attention in the industry and business field due to their biotechnological applications [11]. Lipases have a wide range of uses in industry productions such as diary-based products, detergents, drugs, cosmetics and leather processes. On the other side, Lipase production is a challenging, complex and not easy to model or monitor [10]. The complexity is due to the nature of lipase production which is highly dependent on its operating conditions that affect its growth.

Consequently, deciding upon an optimization method and choosing a modeling technique are vital issues in the

Corresponding Author: Hossam Faris, Business Information Technology Department,
The University of Jordan, Amman, Jordan.

process of producing reliable lipase products with high standards. Efficient optimization and modeling can dramatically improve the system performance and reduce the costs [4]. Accordingly, determining the finest optimization method and choosing a modeling procedure is significant in the process of lipase manufacturing [12,13]. Efficient optimization and modeling can dramatically improve the system performance and reduce the costs of production [4].

In this paper, we explore the use of Multigene Symbolic Regression Genetic Programming [14] to develop a mathematical model which can estimate the Lipase activities in submerged fermentation (SmF). A multigene individual consists of one or more genes, each of which is a "traditional" GP tree [15, 16]. Genes are acquired incrementally byindividuals in order to improve fitness (e.g. to reduce the sum of squared errors on a data set). The overall model is a weighted linear combination of each gene. The proposed Multigene GP model should be able to correctly estimate the lipase activities. A comparison between traditional multiple regression [13] neural networks [17] and fuzzy logic [12] models will be provided. Many models were developed with various setting parameters of GP and the best model was selected. The reported results are promising and can compete with other known soft computing models.

Among artificial intelligence and machine learning approaches,Artificial Neural Networks (ANNs) are the mostapplied in lipase modeling and prediction. Ebrahimpour *et al*. [4] usedthe best composition of production medium among the bestpreviously published media, then they made a comparisonby applying both response surface methodology (RSM) andANN for optimizing the physical factors for extra cellular ther mostable lipase production. Although both techniquesgave good predictions, the ANN showed better performancein data fitting and estimation capabilities. However ANN ingeneral suffers some disadvantages; ANNs relatively needlarge amounts of data for training and they work as blackinput/output box, it is always hard to interpret their results.

Other famous approaches are Genetic Programming (GP)and Genetic Algorithms (GA). Both approaches are evolutionarytechniques inspired from biology concepts. GP and GA applied for modeling lipase production by researches but they are less common than ANN. For example, Ahmed *et al*. [3] applied GP as evolutionary computation methodologyfor developing an efficient model for the fermentation process. The Authors compared their results with other results obtained from traditional experimental design approach (Box-Behnken). Their final results show the superiority of the GP in modeling the fermentation process. Roeva [18], reported a modified genetic algorithm is proposed for a parameter identification of an E. coli fedbatch fermentation model. The authors made some adjustments of the genetic parameters regarding the fermentation processes, to improve the conventional genetic algorithm. Authors claim that the modified GA for a parameter identification of the problem can be efficient and effective. Applying of the modified GA can decrease the running time but relatively still high.

Multiple Regression Approach: This approach uses the method of least squares estimation (LSE), to model a relationship between one dependent and many independent variables. Multiple regression models were used to solve variety of modeling problems. To show how the parameter estimation process work, we assume that a system with i input variables u_i, $i = 1,...,5$ and single output $\hat{y}$ can be modeled with a single function f based a set of n observations. The function f could have a different level of complexity. For simplicity we will assume that f is a simple linear function as given in equation 1. The level of complexity could be higher as we will discusslater in our case study which is given in equation 11. The multiple regression models have the following mathematical representation.

$$\Gamma = f(\zeta)$$

$$= \zeta_0 + \zeta_1 u_1 + \ldots\ldots\ldots + \zeta_5 u_5 + \epsilon_i$$

$$= U_f^T \zeta + \epsilon_i \quad (1)$$

To find the values of the model parameters ζ's we need to build what is called the regressor or exogenous matrix U. This matrix is developed based on the experiment collected measurements. Thus, U can be presented as follows:

$$U = \begin{pmatrix} u_1^1 & \cdots & u_5^1 \\ u_1^2 & \cdots & u_5^2 \\ \vdots & \cdots & \vdots \\ u_1^n & \cdots & u_5^n \end{pmatrix}$$

The parameter vector ζ, the error variable ϵ and the

response variable or dependent variable y can be presented as follows:

$$y = \begin{pmatrix} y^1 \\ y^2 \\ \vdots \\ y^n \end{pmatrix}, \zeta = \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ \zeta_5 \end{pmatrix}, \in = \begin{pmatrix} \in_1 \\ \in_2 \\ \vdots \\ \in_n \end{pmatrix}$$

The least squares solution yields the normal equation: $U\zeta = y$

$$\zeta = \mathrm{U}^{-1}\mathrm{y} \tag{2}$$

If the regression matrix U is not a symmetric matrix, we have to reformulate the equation such that the solution for the parameter vector ζ is as follows:

$$\zeta = (\mathrm{U}^{\mathrm{T}}\mathrm{U})^{-1}\mathrm{U}^{\mathrm{T}}\,\mathrm{y} \tag{3}$$

The second order regression model is given in equation 4. This model can provide a better accuracy than the first order model since it provides more dynamics and non-linearity.

$$y = \zeta_0 + \Sigma_{i=0}^{0}\,\zeta_i\ \mathrm{u} + \Sigma_{i=0}^{n}\Sigma_{j=0}^{n}\,\zeta_{\mathrm{ij}}\mathrm{u_i}\mathrm{u_j} \tag{4}$$

Symbolic regression method was presented by Koza [15]. The objective of this method is to search the space of possible mathematical expressions (i.e. trees) while minimizing some error criteria. Traditional system identification techniques usually adopt two stages of operation: structuredetermination and parameter estimation [19-21]. In each stage, some strategy needs to be adopted to select the suitable class of models and to estimation the model parameters using a technique such as Least Square Estimation (LSE) or Instrumental Variable (IV) method. Symbolic regression, is not similar to traditional linear and nonlinear regression methods, it searches both the space of models along with the space of all possible parameters simultaneously such that it can find the best model which minimize the error criterion.

Multigene Symbolic Regression: Symbolic regression was used in GP to evolve a population of trees [16]. For a system with u input of dimension $R^{n\times n}$ to produce a model output $\hat{y}$ with dimension $R^{n\times 1}$, where n is the number of observations taken and m is the number of input variables, we could produce a tree structure which introduces the mathematical relationship:

$$\hat{y} = f(u_1,...,u_i) \tag{5}$$

In multigene symbolic regression GP, each prediction of the output variable $\hat{y}$ is formed by a weighted output of each of the trees/genes in the multigene individual plus a bias term. Each tree is a function of zero or more of the i input variables $\mathrm{u_1}$, . . . , $\mathrm{u_i}$. Mathematically, a multigene regression GP model can be written as:

$$\hat{y} = y_0 + y_1 \times tree_1 + ... + y_M \times tree_M \tag{6}$$

where y_0 represents the bias or offset term while y_1 , . . . , y_M are the gene weights and M is the number of genes (i.e. trees) which constitute the available individual. The weights (i.e. regression coefficients) are automaticallydetermined by a least squares procedure for each multigene individual.

In multigene symbolic regression each symbolic model is represented by number of GP trees weighted by linear combination. Each tree is considered as a gene by itself. Anexample of multigene model is shown in Figure 1. The given model can be presented mathematically as given in equation 7.

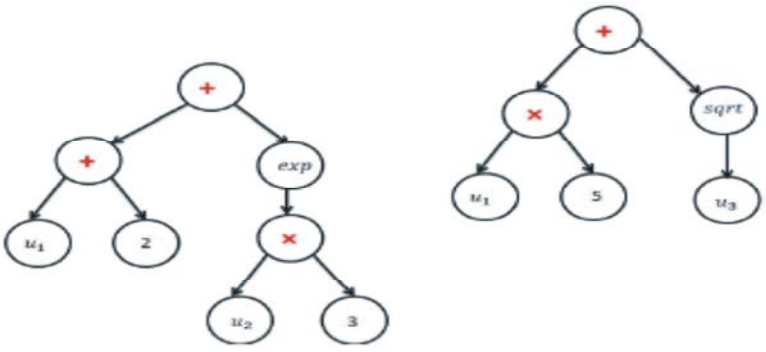

Fig. 1: Example of a multigene symbolic model

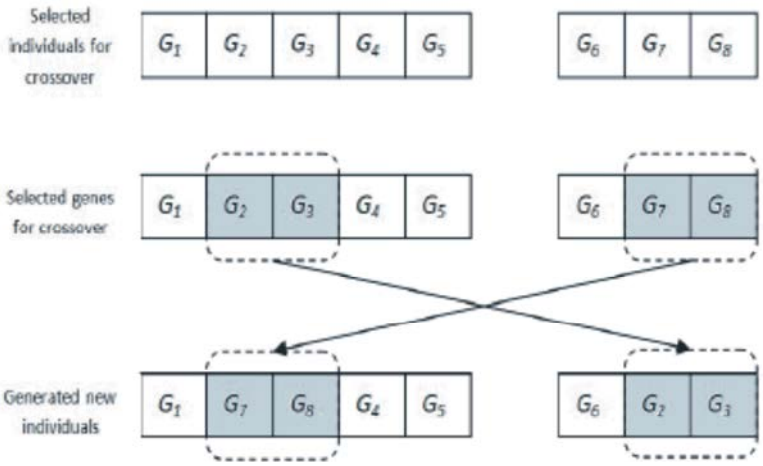

Fig. 2: Example of multigene GP crossover operator

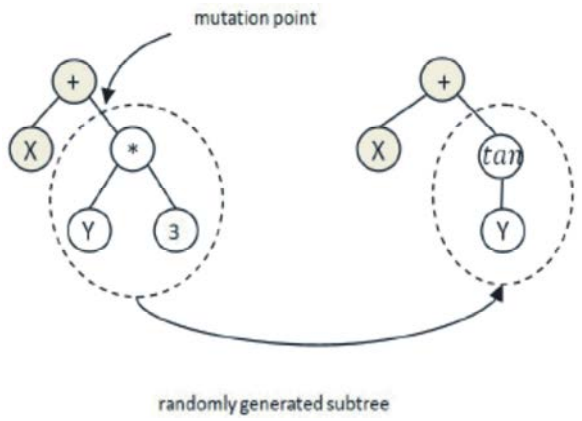

Fig. 3: Example of GP mutation operator

$$y_0 + y_1[(2+u_1) + e^{3\times u\,2}] + y_2[(5u_1) + (\sqrt{u_3})] \quad (7)$$

Initial Population and Representation: The evolutionary cycle of multigene GP starts by generating a number of individuals (candidate models). Each individual consists of randomly generated trees combined with a set of linear coefficients such as y_0 , y_1,.. The maximum number of trees in any given individual equals to G_{max} which can be defined by the user while the linear coefficients are estimated by the ordinary least squares method. Hinchliffe *et al*. [22] reported that the symbolic regression using multigene GP has some advantages over the standard GP. The nature structure of the multigene individuals helps in getting relatively compact and easy to evaluate mathematical models [14].

Terminal and Function Sets Multigene individuals can be defines using a Terminal set *TR* and a Function set F . The set F typically contains arithmetic operators such as addition, subtraction, multiplication and division, also it could contains other non-linear terms such as *sqrt*, *exp*, *sin*, *cos* and more complex functions.

The function set is combined with the terminal set to help the algorithm to develop and form suitable tree structures which represent a model for the problem. This multigene symbolic model has the advantages that it is linear in the parameters with respect to the coefficients γ_0, γ_1 and γ_2.

Tournament Selection: This is one of the most common selection methods in genetic programming. In this method, *s*individuals (i.etournamentMiller *et al*. [23]) are selected randomly from a given generation then the individual which has the best fitness value is chosenfor reproduction.

s is usually referred to as the tournament size. The selection pressure of tournament selection is scalable according to the tournament size.

Crossover: During the evolutionary process, genes are generated using crossover operator. In our case, we used two-point crossover. Crossover makes exchange of genes between individuals. A two point high level crossover adopted by Reza *et al*. [14] is used. The following example in Figure 2 shows the operation.

Mutation: Mutation is an operator applied on a selected single individual. A randomly chosen point in the tree representation of the individual is truncated and replaced with another randomly generated sub tree as shown in Figure 3. The resulted individual replaces the older one. Typically, mutation operator is performed with a probability much less than crossover [24].

Elitism: This operation selects one or more individuals usually based on their fitness value, and copies them to the next generation without any modification [25].

Termination Condition: The evolutionary cycle of the GP algorithm keeps iterating until one of the following conditions is met;

- Maximum number of generations is reached. It is a predetermined number specified by the user to end the iterative process after a number of cycles.
- An individual with a specific fitness value is reached. Finally, the best-so-far individual is chosen to be the solution of the problem.

Model Validation: The performance of the developed Multigene GP modeland the other models used for comparison (i.e., NN, MR and FL) shall be evaluated using the following validation criteria:

- Variance-Accounted-For (VAF):

$$VAF = \left[1 - \frac{\text{var}(y - \hat{y})}{\text{var}(y)}\right] \times 100\% \quad (8)$$

- Euclidian distance (ED):

$$ED = \sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (9)$$

- Manhattan distance (MD):

$$MD = \sum_{i=1}^{n} | y_i - \hat{y}_i | \quad (10)$$

where y and $\hat{y}$ are the actual lipase activates and the estimated activates based on proposed model, respectively. n is the number of observations used in the experiments.

MATERIALS AND METHODS

Bacterial Strain: In this paper, we considered the data set developed from Khoramnia *et al*. [17]. Author explained the methodology of producing the bacterial strain in details. The bacterial strain was isolated from oily food waste in Serdang, Selangor, Malaysia and identified as Acinetobacter sp. by the German Collection of Microorganisms and Cell Cultures (DSMZ), Braunschweig, Germany.

Lipase Production in SmF: Khoramnia*et al.*[17] author stated that:

"The selected SmF lipase production medium was composed of (% w/v): peptone (5), yeast extract (1), NaCl (0.05), CaCl2 (0.05), lactose (1); and coconut oil (1% v/v). The medium was sterilized for 20min at 121C. The SmF cultures were performed in 250mL blue cap bottles in a rotary incubator shaker (0-250rpm). The agitation, inoculum size, initial pH, temperature, and time were adjusted according to the central composite rotatable

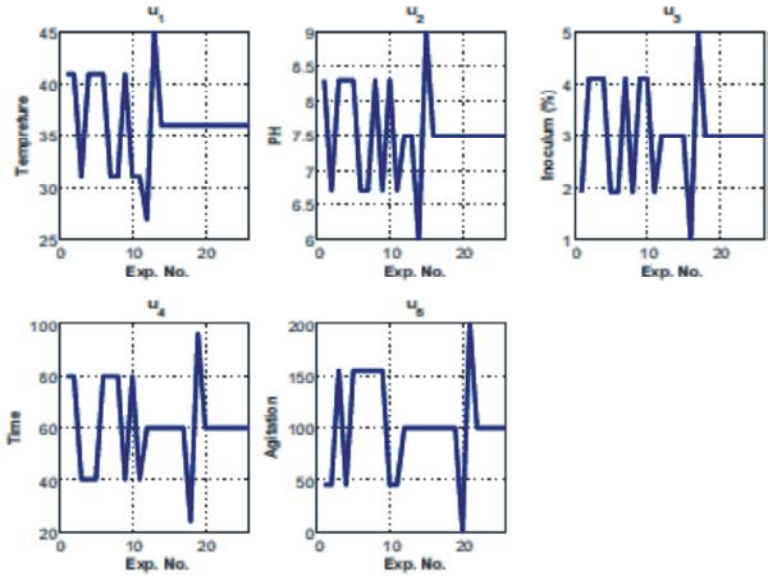

Fig 4: The five independent variables: Temperature, pH, Inoculum, Time and Agitation which contribute to Lipase Activity in SmF

design (CCRD). After lipase production, the cell-free supernatant was obtained by centrifugation at 12,000 x g, 4°C for 10min prior to lipase assay."

Data Collection: Experimental data was collected for five variablesas reported by Khoramnia *et al*. [17] and Cohran and Cox [26] for the lipase production in SmF. The variable selected levels were incubation temperature (27_45°C); initial pH(6-9); moisture content (60-100%); olive oil (0-20%) and incubation period (72-168*h*). Khoramnia *et al*. [17] explored the use of Artificial Neural Networks to model the lipase activities.

The experimental produced lipase activity in SmF is presented in Table IV and shown in Figure 4.

RESULTS AND DISCUSSIONS

Multiple Regression: A multiple regression model was developed to estimate the lipase activities in SmF. The problem of modeling the lipase activities can be viewed as a system identification problem. The model shall represents the relationship between the input variables u_1, u_2, u_3,u_4,u_5 which represent Temperature, pH,Inoculum, Time andAgitation, respectively, and the observed$y$ in a certain range of operating conditions. By applying multiple regression analysis on the experimentaldata, the following second-order polynomial equationwas established to explain the lipase production:

$$y = \zeta_0 + \zeta_1 u_1 + \zeta_2 u_2 + \zeta_3 u_3 + \zeta 3\, u_4 + \zeta_5 u_5 + \zeta_6 u_1^2 + \zeta_7 u_2^2 + \zeta_8 u_3^2 + \zeta_9 u_4^2 + \zeta_{10} u_5^2 + \zeta_{11} u_1 u_2 + \zeta_{12} u_1 u_3 + \zeta_{13} u_1 u_4 + \zeta_{14} u_1 u_5 + \zeta_{15} u_2 u_3 + \zeta_{16} u_2 u_4 + \zeta_{17} u_2 u_5 + \zeta_{18} u_3 u_4 + \zeta_{19} u_3 u_5 + \zeta_{20} u_4 u_5 \quad (11)$$

where $\hat{y}$ is the predicted lipase activities, ζ_0 model constant;u_1 , u_2 , u_3 , u_4 and $u_5$are independent variables; ζ_1 ,ζ_2 , ζ_3 , ζ_4 and ζ_5 are linear coefficients; ζ_6 , ζ_7 , ζ_8 , ζ_9 and $\zeta_{10}$are the quadratic coefficients; ζ_{11} , ζ_{12} , ζ_{13} , . . . are the crossproduct coefficients. The values of the parameters ζ wereobtained by solving this regression problem. We estimatedthe parameters of this model using Least Square Estimation (LSE). The model parameters are given in Table I. Figure 5 show the actual and estimated lipase activities using MR model.

Genetic Programming

GPTIPS Programming Tool: To develop our proposed Multigene GP model, we used the GPTIPS MATLAB Toolbox developed in Reza *et al*. [14]. GPTIPS is a genetic programming software tool which can be used for modeling dynamical nonlinear systems. The tool can be configured to evolve multigene tree structure. In GPTIPS, the optimal weights for the genes are automatically obtained using ordinary least squares to regress the genes against the output data. The resulting pseudo-linear model can capture non-linear behavior. The multigene approach often develops simpler models than evolving models consisting of one monolithic GP tree [14].

The number and structure of the trees is evolved automatically during each run. The GP base software can be used to define the number of trees to be combined. As the number of tress increased the model complexity increased but a possible solution could be found. Training data (i.e. input/output measurements) are used to develop the model. Testing data are used, after the run, to evaluate the developed (i.e. evolved) models.

GP Setup: Some parameters have to be defined by the user at the beginning of the evolutionary process. They include: population size, probability of crossover, mutation probability and the type of the selection mechanism. User has also to setup the maximum number of genes G max a model is allowed to have. The maximum tree depth D max allows us to change the complexity of the evolved models. Restricting the tree depth helps evolving simple model but it may also reduce the performance of the evolved model. A prior knowledge on

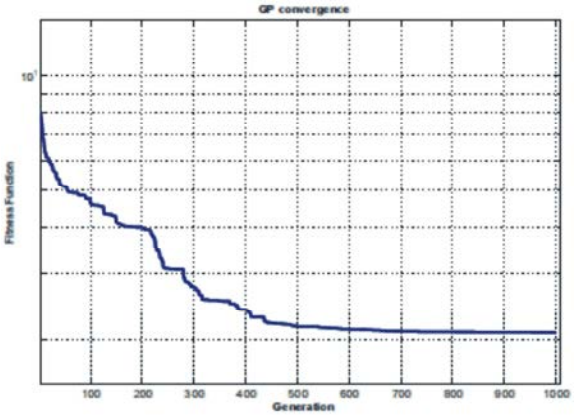

Fig. 5: Convergence of GP over 1000 generations

Table 1: Estmaited values of ζ 's using lse with ζ 0 =155.6888

ζ_1	ζ_2	ζ_3	ζ_4	ζ_5
-1.6136	-22.8679	-20.4559	-0.7810	0.3274
ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
0.0028	1.0607	-0.4272	-0.0010	-0.0003
ζ_{12}	ζ_{13}	ζ_{14}	ζ_{15}	ζ_{16}
0.0570	-0.0119	0.0599	0.0195	-0.0003

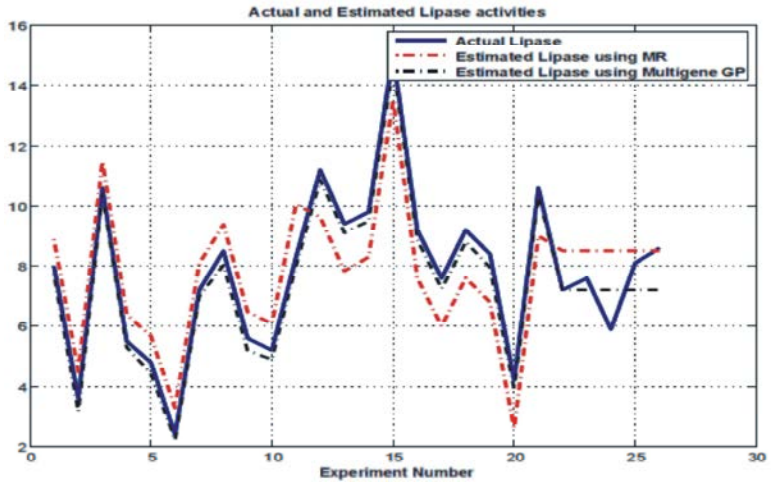

Fig. 6: Actual and Estimated Lipase Activities Model: A Comparison between MR and Multigene GP

Table 2: GP TUNING PARAMETERS

Population size	50
Number of generation	1000
Selection mechanism	Tournament
Max. tree depth	12
Max. No. of genes allowed in an individual	7
Elite	0.2

Table 3:Evaluation criteria for the developed models

NO.	ANN model[17]	MR model [12]	FL model [13]	GP model
VAF	64.15%	79.26%	90.98%	97.58%
ED	3.90	3.89	2.18	2.61
MD	27.50	27.63	15.34	0.41

Table 4: Actual and estimated lipase activities in smf based models inspired by NN, MR, FL and multigene GP techniques

Temp. C°	pH	Inoculum (%)	Time (h)	Agitation (rpm)	Lipase Act. Measured	NN Model [17]	MR Model [12]	FL Model [13]	GP Model
40.9	8.3	1.9	79.8	45.1	8	7.9	8.8808	7.3466	7.7227
40.9	6.7	4.1	79.8	45.1	3.6	3.7	4.4808	2.3755	3.1707
31.1	8.3	4.1	40.2	154.9	10.6	7	11.4808	9.9609	10.4069
40.9	8.3	4.1	40.2	45.1	5.5	7	6.3808	7.3989	5.2982
40.9	8.3	1.9	40.2	154.9	4.8	4.1	5.6808	5.1985	4.4379
40.9	6.7	1.9	79.8	154.9	2.4	1.1	3.2808	2.4247	2.2225
31.1	6.7	4.1	79.8	154.9	7.2	7.5	8.0808	6.5098	6.9982
31.1	8.3	1.9	79.8	154.9	8.5	6.2	9.3808	10.0611	8.0316
40.9	6.7	4.1	40.2	154.9	5.6	6.3	6.4808	5.7953	5.1902
31.1	8.3	4.1	79.8	45.1	5.2	5.6	6.0808	4.7972	4.8965
31.1	6.7	1.9	40.2	45.1	8.3	8.6	10.0617	7.9416	8.0151
27	7.5	3	60	100	11.2	10.2	9.6334	11.3319	10.8694
45	7.5	3	60	100	9.4	10.4	7.8334	8.8956	9.1096
36	6	3	60	100	9.8	9.9	8.2967	10.4795	9.4983
36	9	3	60	100	15	13.4	13.4967	13.4307	14.6615
36	7.5	1	60	100	9.2	9.5	7.6013	8.1593	8.8868
36	7.5	5	60	100	7.6	3.4	6.0013	7.5686	7.2467
36	7.5	3	24	100	9.2	11.9	7.6013	8.1315	8.8088
36	7.5	3	96	100	8.4	8.3	6.8013	8.9156	7.9563
36	7.5	3	60	0	4.2	5.8	2.6071	4.5621	3.8975
36	7.5	3	60	200	10.6	9.8	9.0071	10.0889	10.3858
36	7.5	3	60	100	7.2	10	8.5101	8.0652	7.2024
36	7.5	3	60	100	7.6	X	8.5101	8.0652	7.2024
36	7.5	3	60	100	5.9	X	8.5101	8.0652	7.2024
36	7.5	3	60	100	8.1	X	8.5101	8.0652	7.2024
36	7.5	3	60	100	8.6	X	8.5101	8.0652	7.2024

the problem domain could help in designing a function set which could speed up the evolutionary process for model development.

GP Model: The developed genetic programming model output $\hat{y}$ is given in Table IV. Figure 6 show the actual and estimated lipase activities based the developed GP model and MR model. In Figure 5, we show the convergence of GP over 1000 generations. The tuning parameters for GP evolutionary process is given in Table II. In order to compare the results of the genetic programming model with the polynomial one, the VAF was computed also for the multiple regression models. The computed values are given in Table III.

CONCLUSIONS

In this paper a multigene genetic programming approach was used to develop an optimized Multigene GP model the lipase activity. The developed model of this research was based on five independent variables showing observed values of lipase activity in SmF. They are the temperature, pH, inoculum, time and agitation. Multigene genetic programming evolved compact linear combinations of non-linear transformations of the selected input variables. Performance of the developed model was evaluated and compared based on different criteria. Genetic Programming showed promising results compared to other reported approaches used in the literature such as the traditional multiple regression neural networks and fuzzy logic models.

REFERENCES

1. Ghose, T. and V. Bisaria, 2000. Development of biotechnology in india, in History of Modern Biotechnology I (A. Fiechter, ed.), vol. 69ofAdvances in Biochemical Engineering/ Biotechnology, pp: 87-124.
2. Abdel-Fattah, Y., 2002. Optimization of thermostable lipase production from a thermophilicgeobacillus sp. using box-behnken experimental design, Biotechnology Letters, 24: 1217-1222.
3. Sheta, W., W.M. Aly, Y.R. Abdel-Fattah and A.F. Sheta, 2003. Designing a biological experiment using genetic programming: An evolutionary methodology to improve the production of thermostable lipase enzyme, WSEAS Transactions on Systems, 4(2): 1221-1232.
4. Ebrahimpour, A., R. Rahman, D. EanCh´ ang, M. Basri and A. Salleh, 2008. A modeling study by response surface methodology and artificial neural network on culture parameters optimization for thermostable lipase production from a newly isolated thermophilicgeobacillus sp. strain arm, BMC Biotechnology, 8(1): 96.
5. Zamost, B., H. Nielsen and R. Starnes, 1991. Thermostable enzymes for industrial applications, Journal of Industrial Microbiology, 8: 71-81.
6. Olsen, O., R. Borriss, O. Simon and K. Thomsen, 1991. Hybrid bacillus (1- 3,1-4)-?-glucanases: engineering thermostable enzymes by construction of hybrid genes, Molecular and General Genetics MGG, 225: 177-185.
7. Turner, P., G. Mamo and E. Karlsson, 2007. Potential and utilization of thermophiles and thermostable enzymes in biorefining, Microbial Cell Factories, 6: 1-9.
8. Chung, D.H., J. Huddleston, J. Farkas and J. Westpheling, 2011. Identification and characterization of CbeI: a novel thermostable restriction enzyme from caldicellulosiruptorbescii DSM 6725 and a member of a new subfamily of HaeIII-like enzymes," Journal of Industrial Microbiology and Biotechnology, vol. 38, pp. 1867-1877.
9. Schmidt-Dannert, C., M.L. Rua, H. Atomi and R.D. Schmid, 1996. Thermoalkalophilic lipase of bacillus thermocatenulatus. i. molecular cloning, nucleotide sequence, purification and some properties, Biochimica et Biophysica Acta, 1301(1-2): 105114.
10. Cceres, C.X., D.M. Freire, R.E. Cceres and R.F. Segovia, 2005. Mathematical modeling of lipase production by penicilliumrestrictum in batch fermentation, in Proc. of 4th Mercosur Congress on Process Systems Engineering.
11. Emmanuel, A.N., A. David, Y.K. Benjamin and E.T. Yannick, 2009. A hybrid neural network approach for batch fermentation simulation," Australian Journal of basic and Applied Sciences, 3(4): 3930-3936.
12. Hiary, R., A. Sheta and H. Faris, Fermentation process modeling using takagi-sugeno fuzzy model, WSEAS Trans. on Systems, 11(8): 375-384.
13. Sheta, A. and R. Hiary, 2012. Modeling lipase production process usingartificial neural networks, in Proceedings of the 3rd IEEE International Conference on Multimedia Computing and Systems, (Tangier, Morocco), pp: 1158-1163, 10-12 May.

14. Searson, D.P., D.E. Leahy and M.J. Willis, 2010. GPTIPS : An open source genetic programming toolbox for multigene symbolicregression, in Proceedings of the International Multi-conference of Engineers and Computer Scientists 2010 (IMECS 2010), vol. 1, (Hong Kong), pp: 77-80, 17-19 Mar.
15. Koza, J., 1991. Evolving a computer program to generate random numbersusing the genetic programming paradigm, in Proceedings of the Fourth International Conference on Genetic Algorithms, Morgan Kaufmann, La Jolla,CA.
16. Koza, J.R., 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. The MIT Press.
17. Khoramnia, A., A. Ebrahimpour, B.K. Beh and O.M. Lai, 2011. Production of a solvent, detergent, and thermotolerant lipase by a newly isolated acinetobacter sp. in submerged and solid-state fermentations, Journal of Biomedicine and Biotechnology, pp: 702179.
18. Roeva, O., 2006. A modified genetic algorithm for a parameter identification of fermentation processes, Biotechnology and Biotechnological Equipment, 20(1): 202-209.
19. Iba, H., T. Hitoshi, G. Hung and S. Taisuke, 1993. System identification using structured genetic algorithms, in Proceedings of the FifthInternational Conference on Genetic Algorithms, pp: 279-286, Morgan Kaufmann.
20. Sheta, A., K.D. Jong, J. Gertler and O. Frieder, 1997. System identification using hybrid genetic algorithms, in Proceedings of the IEEE International Conference on Electronics, Circuits and Systems (ICECS'97), Cairo, Egypt.
21. Chen, S., S.A. Billings and P.M. Grant, 1990. Non-linear system iden- tification using neural networks, International J. Control, 51: 1191-1214.
22. Hinchliffe, M., H. Hiden, B. McKay, M. Willis, M. Tham and G. Barton, 1996. Modelling chemical process systems using a multi-gene genetic programming algorithm, in Late Breaking Papers at the Genetic Programming 1996 Conference Stanford University July 28- 31, 1996 (J. R. Koza, ed.), (Stanford University, CA, USA), pp. 56-65, Stanford Bookstore, 28-31 July.
23. Miller, B.L. and D.E. Goldberg, 1995. Genetic algorithms, tournament selection, and the effects of noise, Complex Systems, 9: 193-212.
24. Keller, R.E. and W. Banzhaf, 1996. Genetic programming using mutation, reproduction and genotype-phenotype mapping from linear binarygenomes into linear lalr(1) phenotypes, in in Proceedings of Genetic Programming 1996 Conference, pp: 116-122, MIT Press.
25. Yang, S., 2008. Genetic algorithms with memory and elitism based immigrants in dynamic environments, Evolutionary Computation, pp: 385- 416.
26. Cohran, W.G. and G.M. Cox, 2002. Experimental Design. John Wiley and Sons, New York, NY, USA.