**Research Article**

# A Novel LSSVM Based Algorithm to Increase Accuracy of Bacterial Growth Modeling

**Masoud Salehi Borujeni [1†], Mostafa Ghaderi-Zefrehei [2*], Farzan Ghanegolmohammadi [3†] and Saeid Ansari - Mahyari [4]**

[1] Electronics Department, Electrical and Computer Engineering Faculty, Semnan University, Semnan, Iran

[2] Department of Animal Science, Faculty of Agriculture, Yasouj University, Yasouj, Iran

[3] Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba, Japan

[4] Department of Animal Science, College of Agriculture, Isfahan University of Technology, Isfahan, Iran

*Corresponding author*: Mostafa Ghaderi-Zefrehei, Department of Animal Science, Faculty of Agriculture, Yasouj University, Pasdaran Ave., Yasouj, Iran, 75914, Telefax: +98 741 222 4840, E-mail: mghaderi@yu.ac.ir.

[†] Masoud Salehi Borujeni and Farzan Ghanegolmohammadi contributed equally to this work.

**Background:** The recent progress and achievements in the advanced, accurate, and rigorously evaluated algorithms has revolutionized different aspects of the predictive microbiology including bacterial growth.

**Objectives:** In this study, attempts were made to develop a more accurate hybrid algorithm for predicting the bacterial growth curve which can also be applicable in predictive microbiology studies.

**Materials and Methods:** Sigmoid functions, including Logistic and Gompertz, as well as least square support vector machine (LSSVM) based algorithms were employed to model the bacterial growth of the two important strains comprising *Listeria monocytogenes* and *Escherichia coli*. Even though cross-validation is generally used for tuning the parameters in LSSVM, in this study, parameters tuning (*i.e.*, '*c*' and '$\sigma$') of the LSSVM were optimized using non-dominated sorting genetic algorithm-II (NSGA-II), named as NSGA-II-LSSVM. Then, the results of each approach were compared with the mean absolute error (MAE) as well as the mean absolute percentage error (MAPE).

**Results:** Applying LSSVM, it was resulted in a precise bacterial growth modeling compared to the sigmoid functions. Moreover, our results have indicated that NSGA-II-LSSVM was more accurate in terms of prediction than LSSVM method.

**Conclusion:** Application of the NSGA-II-LSSVM hybrid algorithm to predict precise values of '*c*' and '$\sigma$' parameters in the bacterial growth modeling resulted in a better growth prediction. In fact, the power of NSGA-II for estimating optimal coefficients led to a better disclosure of the predictive potential of the LSSVM.

*Keywords:* Bacterial growth curve, Hybrid algorithm, LSSVM, NSGA-II, Modeling

## 1. Background

Prediction is a part of considerable numbers of scientific studies (1, 2). Also, time series analysis is an active research area in the modeling of the growth curves (1, 3-5). However, the accuracy of the applied model is a critical issue for many decisions making processes (2, 5). The advent of powerful computational approaches has enabled predictive microbiology to anticipate the behavior of a wide range of microorganisms under various environmental stimuli (6, 7). The complicated modifications in predictive microbiology models such as models describing individual cell behavior (8), individual lag time (9, 10), secondary models considering environmental factors, and their interactions (11, 12) have been developed for an accurate prediction of the experimental growth curves under different conditions (6, 7, 13).

Bacterial growth contains four stages, including *lag* phase, *log* or exponential phase, stationary phase, and death phase. Each phase has computationally

been investigated (14-16) such as comparison of the bacterial growth parameters (17), formulation of the transition from stationary to exponential phases (18), estimation of *lag* phase duration (14), approximation of the maximum growth rate (19), predictive models of the growth curve, *etc.* (for review, see 20).

Least square support vector machine (LSSVM) is derived from the Support Vector Machine (SVM) approach, as described previously (20), providing immense computational advantages over the general standard SVM. Through combination of a sophisticated mathematical approaches in a single bundle, LSSVM is a cutting edge approach between theory and the real-world problems which is the advantage of the machine learning methods over that of statistical studies in various fields of science such as clinical trials (21, 22), financial studies (23), engineering (1, 24, 25), *etc.* Nonetheless, the potentiality of the LSSVM approaches have been neglected in the bacterial research, and a few studies have been published, such as prediction of the adaptive colony segmentation (26), ecotoxicity of the ionic liquids (27), and estimation of the *Escherichia coli* promoter gene sequences (22).

The estimation of the parameters to feed the predictive model is a key step in modeling processes. Different optimization algorithms, especially genetic algorithm (GA), have widely been employed for parameter optimization of the empirical data (2, 28-30). GA is a simple and accurate approach to calculate sigmoid function coefficients. It can also be extended to multi-objective optimization (MOO) method for solving problems with two fitness functions (31).

## 2. Objectives

Even though growth predictive models are specifically developed to provide conservative predictions of a given microorganism's growth under various conditions (11, 12), validation studies of the published models are not precise enough, which could be due to the sensitive nature of the applied algorithms. Considering the predictive potential of the LSSVM in a functional prediction of experimental data, here, LSSVM based algorithms are addressed to the model in a more accurate bacterial growth curve. In this way, a novel so-called non-dominated Sorting Genetic Algorithm-II (NSGA-II)-LSSVM algorithm has been employed for learning LSSVM tuning parameters. Then, the general performance of the proposed algorithm for both train and test datasets has been compared with GA-LSSVM, simplex-LSSVM, and sigmoid functions including Logistic and Gompertz. This comparison has been carried out using the mean absolute error (MAE)

and the mean absolute percentage error (MAPE). All algorithms were fitted to the two important bacterial strains *Listeria monocytogenes*, an important food-borne pathogen (6, 13) and *Escherichia coli*, a renowned prokaryotic model organism in a vast range of biological studies.

## 3. Materials and Methods

Bacterial growth was measured as *log* colony formation unite cfu.mL$^{-1}$ (18). Growth data of the two *L. monocytogenes* datasets with different initial population size (**Supplementary Table 1**a and b), over a period of 552 h were obtained from Augustin *et al.* (11). Furthermore, the growth data of the *E. coli* was profiled by the optical density (OD) using spectrophotometry at 610 nm. For *E. coli* dataset, the following information was collected over an experimental period of 13 h. *E. coli* was batch cultured in a liquid LB (Luria-Bertani) medium at 37 ˚C and the OD was measured (every 30 min) until the death phase (Supplementary Table 1c).

### 3.1. Growth Models

Two paradigms were used to model bacterial growth data: sigmoid functions (Eq. 1-3) and LSSVM based algorithms (Eqs. 4-18).

### 3.1.1. Sigmoid Functions Formulation

Sigmoid functions are the general ways of computationally treating bacterial growth data.

The logistic function is defined as follow (Eq. 1 and 3):

$$y = \cfrac{A}{\left\{1 + exp\left[\cfrac{4.\mu_m}{A}(\lambda - t) + 2\right]\right\}} \qquad (1)$$

Gompertz function is developed as follow (Eq. 2 and 3):

$$y = A\, exp\left\{-exp\left[\cfrac{\mu_m e}{A}(\lambda - t) + 1\right]\right\} \qquad (2)$$

Where
$$y = \ln\left(\cfrac{N(t)}{N_0}\right) \qquad (3)$$

Whereas, '$\mu m$' represents the maximum living bacteria in a batch culture, '$\lambda$' stands for the *lag* growth phase of the bacteria, which is in general small. '$A$' indicates the death phase, $e=exp(1)$, '$N_0$' shows initial bacterial population, and '$N(t)$' is the bacterial population at '$t$' time (19) (for comprehensive information see 30). Coefficients of the Logistic and Gompertz functions

were calculated via trust-region optimization method (32) in curve fitting toolbox of MATLAB (version 8.5.0.197613 - R2015a).

### 3.1.2. LSSVM Mathematical Formulation

The pivotal difference between SVM and LSSVM is that LSSVM solves linear systems instead of quadratic programming. In LSSVM for function estimation (*i.e.*, here prediction of the bacterial growth curve), the optimization problem is formulated (20) as:

$$\min J(w,\varepsilon) = \frac{1}{2}\|w\|^2 + \frac{1}{2}c\sum_{i=1}^{N}\varepsilon_i^2 \qquad (4)$$

Subject to the equality constraints:

$$y_i = w^T\Phi(x_i) + b + \varepsilon_i \qquad (5)$$

Where, '$\Phi(x)$' is a nonlinear function that maps input space into a high dimensional space, '$w$' is weighting vector and '$b$' is bias, '$c$' is the regularization (tuning) parameter determining the trade-off between the training error minimization and smoothness, and '$\varepsilon$' is the training error. The corresponding Lagrange function can be written as:

$$L(w,b,\varepsilon,a) = \frac{1}{2}\|w\|^2 + \frac{1}{2}c\sum_{i=1}^{N}\varepsilon_i^2 - \sum_{i=1}^{N}a_i\left[w^T\Phi(x_i) + b + \varepsilon_i - y_i\right] \qquad (6)$$

In which, '$a_i$' is Lagrange multipliers (or support vectors). Taking the partial derivatives with respect to '$w$','$b$','$\varepsilon$','$a$', and equating them to zero, the conditions for optimality are obtained as follows:

$$\begin{cases} \dfrac{\partial L(w,b,\varepsilon,a)}{\partial w} = 0 \rightarrow w = \sum_{j=1}^{N}a_j\Phi(x_j) \\[2mm] \dfrac{\partial L(w,b,\varepsilon,a)}{\partial b} = 0 \rightarrow \sum_{j=1}^{N}a_j = 0 \\[2mm] \dfrac{\partial L(w,b,\varepsilon,a)}{\partial e_j} = 0 \rightarrow a_j = ce_j \quad j=1,\ldots,N \\[2mm] \dfrac{\partial L(w,b,\varepsilon,a)}{\partial a_j} = 0 \rightarrow a_j = w^T\Phi(x_j) + b + e_j \quad j=1,\ldots,N \end{cases} \qquad (7)$$

These conditions lead to the following linear system:

$$\begin{bmatrix} I & 0 & 0 & -Z^T \\ 0 & 0 & 0 & -Y^T \\ 0 & 0 & \gamma^I & -I \\ Z & Y & I & 0 \end{bmatrix}\begin{bmatrix} w \\ b \\ e \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vec{1} \end{bmatrix} \qquad (8)$$

Where

$$Z = \left[\varphi(x_1)^T y_1; \ldots; \varphi(x_n)^T y_n\right] \qquad (9)$$

$$Y = \left[y_1; \ldots; y_n\right] \qquad (10)$$

$$1 = \left[1; \ldots; 1\right] \qquad (11)$$

$$e = \left[e_1; \ldots; e_n\right] \qquad (12)$$

$$a = \left[a_1; \ldots; a_n\right] \qquad (13)$$

The solution can be written as:

$$\begin{bmatrix} 0 & -Y^T \\ Y & ZZ^T + \gamma I \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ \vec{1} \end{bmatrix} \qquad (14)$$

$$\Omega = ZZ^T \qquad (15)$$

Using Mercer's condition:

$$\Omega_{il} = y_i y_l \varphi(x_i)^T \varphi(x_l) = y_i y_l k(x_i, x_l) \qquad (16)$$

Therefore, the fitting LSSVM regression would be:

$$y(x) = \sum_{i=1}^{N}a_i K(x, x_i) + b \qquad (17)$$

Several choices are possible for the kernel $K(.,.)$. Some typical choices are linear LSSVM (20): $K(x,x_i) = x_i^T x$, polynomial LSSVM (20) of the degree $d$: $K(x,x_i) = (t + x_i^T x)^d$ and radial basis function (RBF) kernel (20): $K(x,x_i) = \exp\left(\frac{-0.5x - x_i^2}{\sigma^2}\right)$. RBF is a common and powerful function to solve regression problems (20, 21, 33). In fact, the RBF kernel, unlike the linear kernel, nonlinearly maps samples into a higher dimensional space; therefore, handles the cases when the relation between inputs and outputs vectors is nonlinear. Furthermore, in comparison to the polynomial kernel, the RBF kernel has less number of hyperparameters which results in the fewer numerical difficulties (see 34 for further information).

In order to use LSSVM model, '$a_i$' and '$b$' coefficients should be calculated. These coefficients are calculated using linear equations of Eq. 8. In this case, first of all, '$c$' and '$\sigma$' coefficients should be determined. Determination of proper coefficients is critical for prediction accuracy.

### 3.1.3. N-fold Cross-Validation

N-fold cross-validation is a common method for calculating '$c$' and '$\sigma$' coefficients. This approach starts
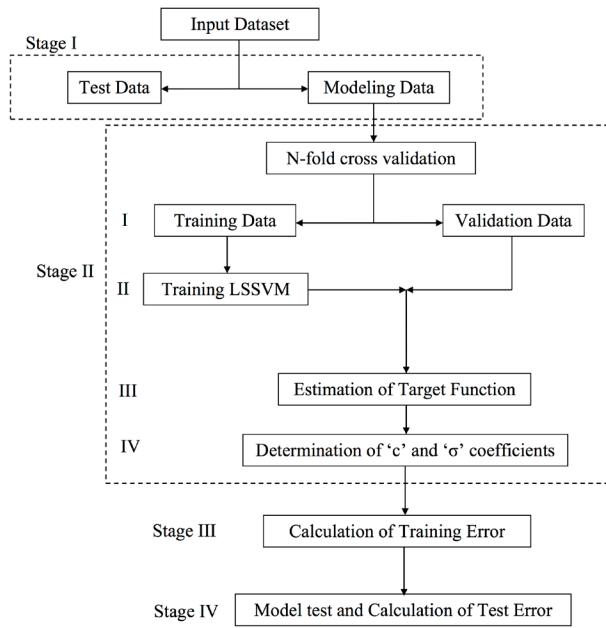
**Figure 1**: Flowchart of N-fold cross-validation method. **Stage I:** Dividing dataset into two sets: modeling (80 %) and test (20 %). **Stage II:** Calculation of 'c' and 'σ' coefficients via GA or simplex algorithms; **II-I:** Dividing dataset into two sets of training and validation through cross-validation method; **II-II:** Training of LSSVM per different 'c' and 'σ' coefficients; **II-III:** Test of trained LSSVM model and fitness function calculation (Eq. 18); **II-IV:** Iteration of stages **II-I, II-II,** and **II-III** for obtaining the best 'c' and 'σ' coefficients. **Stage III:** Training of LSSVM model using modeling dataset and calculation of modeling error. **Stage IV:** Test of LSSVM model using test dataset, and calculation of test error.

with the dividing data into modeling (80%) and test (20%) datasets (**Fig. 1** - Stage I). The modeling dataset is used to estimate 'c' and 'σ' coefficients via GA or simplex algorithms, and the test dataset is used to test the accuracy of the final trained LSSVM model. Then, the modeling dataset is randomly divided into different groups (5 groups in this study); in every iteration, one group is considered as validation and the others as a training dataset. Afterward, LSSVM was trained per different 'c' and 'σ' coefficients which were determined through an optimization algorithm. Next, the trained LSSVM model was tested using validation data, and finally, fitness function (Eq. 18) was calculated (Fig. 1 - Stage II-I to IV). This cycle is terminated by finding the minimum fitness function. Lastly, the obtained 'c' and 'σ' coefficients are the final optimized coefficients (35).

$$ F = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\left| G_{validation}\left(i,j\right) - G_{real}\left(i,j\right) \right|}{G_{real}\left(i,j\right)} \tag{18} $$

Where, 'n' is the number of groups, 'm' represents number of data in each group, '$G_{validation}(i,j)$' is calculated value from '$j^{th}$' data of '$i^{th}$' group via LSSVM model, and '$G_{real}(i,j)$' is '$j^{th}$' data of '$i^{th}$' group in experimental data.

Generally, 'c' and 'σ' coefficients are optimized using simplex-LSSVM. In this paper, GA has been additionally used to calculate 'c' and 'σ' coefficients as well as the fitness function. Then, modeling error and test error were calculated through modeling dataset, trained LSSVM (stage III), and test dataset (stage IV), respectively.

*3.1.4.  Proposed Method*
Our novel hybrid method; the so-called NSGA-II-LSSVM, was also applied in the present study. The optimization via NSGA-II-LSSVM begins with dividing datasets into modeling and test datasets. Then, the test dataset was used to test the results of modeling dataset. Also, in order to conserve model's smoothness and improving learning accuracy, the modeling dataset was divided into training and validation datasets. In our method, LSSVM was trained using the training dataset, and tested using the validation data. Then, training error was considered as the first fitness function (*f1*) (Eq. 19), and validation error as the second fitness function (*f2*) (Eq. 20) (**Fig. 2**).

We speculate that minimization of the two fitness functions (Eq. 19 and 20) will result in the calculation of 'c' and 'σ' coefficients, and acquired coefficients will provide a more accurate and smooth LSSVM modeling of the bacterial growth curve.

$$ f1 = \frac{1}{n} \sum_{i=1}^{n} \frac{\left| G_{train}\left(t\right) - G_{t_{real}}\left(t\right) \right|}{\left| G_{t_{real}}\left(t\right) \right|} \tag{19} $$

Where 'n' is the number of training data, '$G_{train}(t)$' stands for logarithmic bacterial grow at 't' time, and '$G_{t_{real}}\left(t\right)$' show the original value of growth in logarithmic scale at 't' time.

$$ f2 = \frac{1}{m} \sum_{i=1}^{m} \frac{\left| G_{validation}\left(t\right) - G_{v_{real}}\left(t\right) \right|}{\left| G_{v_{real}}\left(t\right) \right|} \tag{20} $$

Where 'm' is the number of validation data, '$G_{validation}(t)$' represents logarithmic bacterial growth at 't' time, and '$G_{v_{real}}\left(t\right)$' is the original amount of bacterial growth in logarithmic scale at 't' time.

In order to calculate 'c' and 'σ' coefficients, NSGA-II: A multi-objective optimization algorithm toolbox
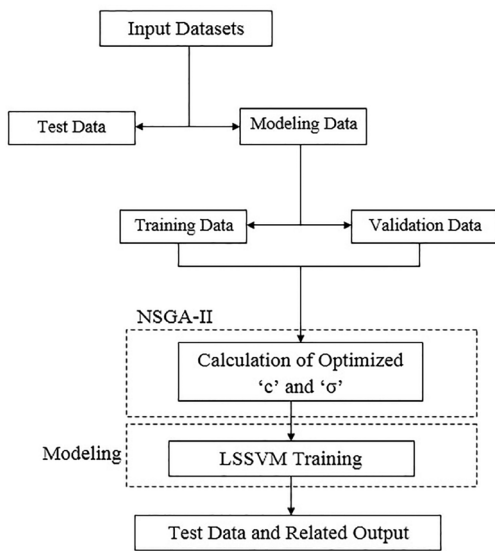
**Figure 2.** The flowchart of data optimization through NSGA-II in LSSVM. Generation of the modeling and test datasets. The obtained test datasets were utilized to test the trained model. The modeling datasets were also divided into training and validation datasets. Training and validating of LSSVM were done using training and validation datasets, respectively. Then, fitness function was defined as training error and validation error. Finally, 'c' and 'σ' coefficients were calculated by minimization of the two obtained fitness functions.

(http://www.mathworks.com) was employed. For complementary information about NSGA-II see **Supplementary Note 1**.

### 3.2. Comparison of the Results
In order to compare the accuracy of the employed algorithms, mean absolute error (MAE) (Eq.21) and mean absolute percentage error (MAPE %) (Eq.22) were used (36).

$$MAE = \frac{\sum_{i=1}^{n} \left| G(t) - G_{real}(t) \right|}{n} \qquad (21)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{\left| G(t) - G_{real}(t) \right|}{G_{real}(t)} \times 100 \qquad (22)$$

Where, '$G(t)$' is the logarithmic bacterial growth at time trained by model, and '$G_{real}(t)$' is real logarithmic bacterial growth at '$t$' time.

MATLAB (version 8.5.0.197613 - R2015a) codes of implementation of LSSVM using the optimized '$c$' and '$σ$' coefficients obtained from NSGA-II, for *E. coli* growth dataset are presented as **Supplementary Note 2**.

## 4. Results
Coefficients of Logistic and Gompertz functions for each dataset are presented in **Tables 1** and **2**. Moreover, '$c$' and '$σ$' coefficients of LSSVM based models are shown in **Table 3**. It is important to be stated that '$c$' values should be greater than zero.

Simplex-LSSVM has displayed the more accurate result in comparison with sigmoid functions (**Figs. 3-5, Supplementary File 1** and **Table 4**). In the first dataset, comparing simplex-LSSVM with Logistics and Gompertz functions, MAPE of modeling indicated an improvement in the accuracy of the prediction using simplex-LSSVM (1.601, 5.345 and 3.984 for simplex-LSSVM, Logistic and Gomperetz models, respectively). Consistent results were also observed in MAPE of testing dataset (Table 4). Moreover, the obtained analogous results regarding the other two datasets and MAE value (Table 4) indicated that LSSVM is obviously superior to the sigmoid functions. Comparing different LSSVM based algorithms, using both MAPE and MAE measures, it was revealed that even though GA-LSSVM and NSGA-II-LSSVM are superior to simplex-LSSVM, NSGA-II-LSSVM is a

**Table 1.** Estimated parameters of Logistics function using curve fitting method.

| Dataset | A[1] | $\mu_m$[2] | $\lambda$[3] | Ref. |
|---|---|---|---|---|
| *Listeria monocytogenes* (Dataset No. 1) | 14.2 | 96.68 | 0.059 | (11) |
| *Listeria monocytogenes* (Dataset No. 2) | 10.15 | 86.25 | 0.059 | (11) |
| *Escherichia coli* (Dataset No. 3) | 3.26 | 1.06 | 1.33 | This study |

[1]Death phase
[2]Maximum alive bacteria in the batch culture (1/h)
[3]*Lag* growth phase (h)

**Table 2.** Estimated parameters of Gompertz function using curve fitting method.

| Dataset | A[1] | $\mu_m$[2] | $\lambda$[3] | Ref. |
|---|---|---|---|---|
| *Listeria monocytogenes* (Dataset No. 1) | 14.91 | 0.054 | 78.39 | (11) |
| *Listeria monocytogenes* (Dataset No. 2) | 10.55 | 0.053 | 72.85 | (11) |
| *Escherichia coli* (Dataset No. 3) | 3.306 | 1.005 | 1.16 | This study |

[1]Death phase
[2]Maximum alive bacteria in the batch culture (1/h)
[3]*Lag* growth phase (h)

**Table 3.** LSSVM parameters over Simplex, GA and NSGAII optimization algorithms.

| Algorithms | Simplex | | GA | | NSGA-II | | Ref. |
|---|---|---|---|---|---|---|---|
| **Dataset** | c | σ | c | σ | c | σ | |
| *Listeria monocytogenes* (Dataset No. 1) | 1542.1 | 0.71 | 390.7 | 0.52 | 449.9 | 0.36 | (11) |
| *Listeria monocytogenes* (Dataset No. 2) | 2016.4 | 0.99 | 318.5 | 0.53 | 912.3 | 0.62 | (11) |
| *Escherichia coli* (Dataset No. 3) | 860851.5 | 0.65 | 3889.9 | 0.41 | 117453.1 | 0.50 | This study |
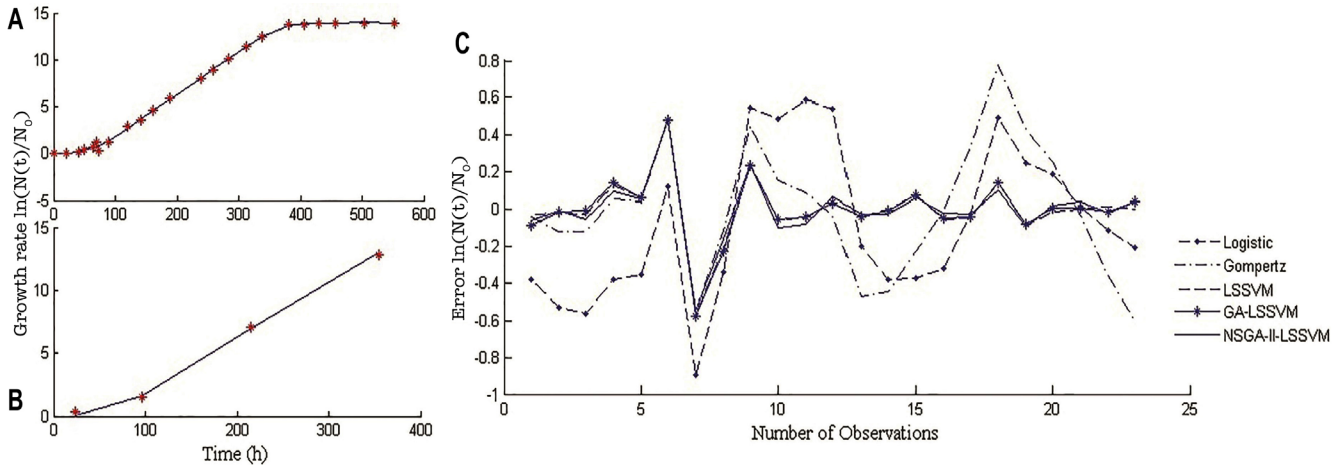


**Figure 3.** NSGA-II-LSSVM hybrid algorithm for *L. monocytogenes* (**Dataset No. 1**) growth modeling. 'N0' displays initial bacterial population, and 'N(t)' is bacterial population at 't' time. **A:** Modeling data, **B:** Test data, accuracy of the applied algorithm is defined by the proximity between predicted and observed values, **C:** Comparing modeling errors of the applied approaches including sigmoid functions (Logistic and Gompertz) and LSSVM based algorithms (Simplex, GA-LSSVM and NSGA-II-LSSVM) (Supplementary File 1).
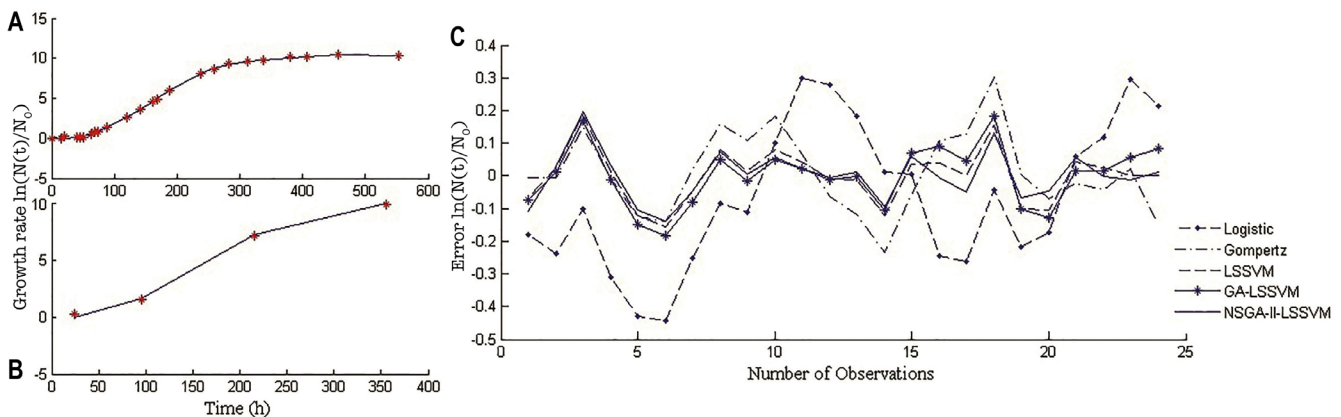


**Figure 4.** NSGA-II-LSSVM hybrid algorithm for *L. monocytogenes* (**Dataset No. 2**) growth modeling. 'N0' displays the initial bacterial population, and 'N(t)' is bacterial population at the 't' time. **A:** Modeling data, **B:** Test data, accuracy of the applied algorithm is defined by the proximity between predicted and the observed values, **C:** Comparing modeling errors of the applied approaches including sigmoid functions (Logistic and Gompertz) and LSSVM based algorithms (Simplex, GA-LSSVM and NSGA-II-LSSVM) (Supplementary File 1).
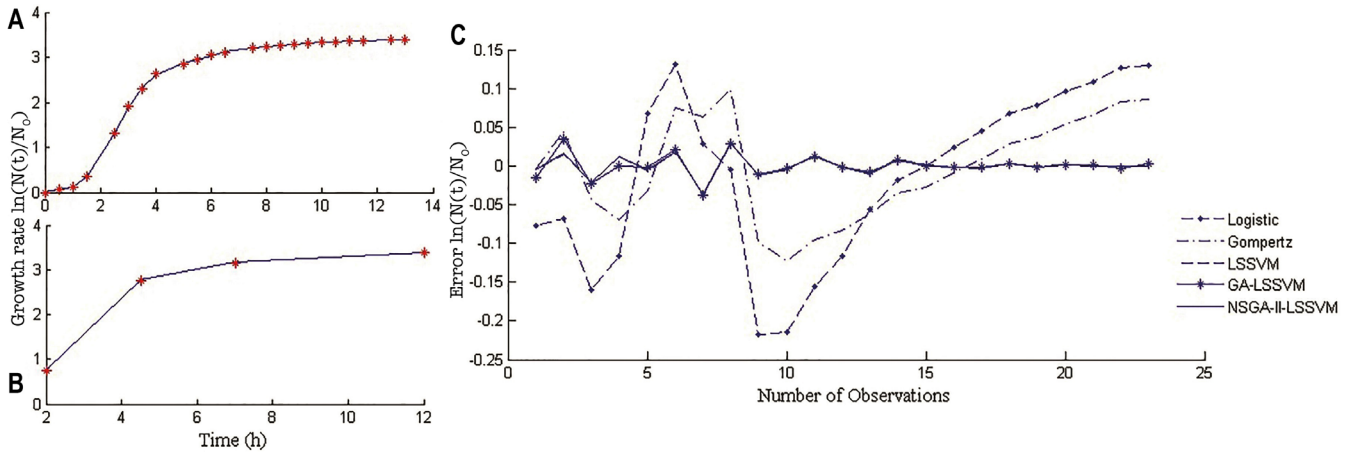
**Figure 5.** NSGA-II-LSSVM hybrid algorithm for *E. coli* (**Dataset No. 3**) growth modeling. 'N0' displays the initial bacterial population and 'N(t)' is bacterial population at 't' time. **A:** Modeling data, **B:** Test data; accuracy of the applied algorithm is defined by the proximity between the predicted and the observed values, **C:** Comparing modeling errors of the applied approaches including sigmoid functions (Logistic and Gompertz) and LSSVM based algorithms (Simplex, GA-LSSVM and NSGA-II-LSSVM) (Supplementary File 1).

**Table 4.** The result of sigmoid and LSSVM based algorithms over three bacterial datasets.

| Model | *Listeria monocytogenes* (Dataset No. 1) | | | | *Listeria monocytogenes* (Dataset No. 2) | | | | *Escherichia coli* (Dataset No. 3) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ref.: | (11) | | | | (11) | | | | This study | | | |
| | Modeling | | Test | | Modeling | | Test | | Modeling | | Test | |
| | MAPE[1] % | MAE[2] | MAPE % | MAE | MAPE % | MAE | MAPE % | MAE | MAPE % | MAE | MAPE % | MAE |
| Logistic | 5.345 | 0.360 | 4.554 | 0.239 | 4.142 | 0.194 | 3.746 | 0.175 | 3.686 | 0.091 | 4.37 | 0.109 |
| Gompertz | 3.984 | 0.268 | 3.818 | 0.208 | 2.036 | 0.095 | 2.821 | 0.131 | 2.318 | 0.057 | 3.070 | 0.076 |
| Simplex-LSSVM | 1.601 | 0.106 | 3.142 | 0.169 | 1.312 | 0.061 | 2.418 | 0.111 | 0.401 | 0.010 | 0.748 | 0.018 |
| GA-LSSVM | 1.579 | 0.106 | 3.110 | 0.169 | 1.218 | 0.059 | 2.402 | 0.109 | 0.391 | 0.009 | 0.748 | 0.018 |
| NSGAII-LSSVM | 1.565 | 0.105 | 3.091 | 0.168 | 1.214 | 0.056 | 2.346 | 0.109 | 0.358 | 0.008 | 0.722 | 0.018 |

[1] Mean Absolute Percentage Error

[2] Mean Absolute Error

more advantageous approach for the modeling bacterial growth curve (less MAPE and MAE values) (Table 4). However, no difference was observed in MAE of the test dataset of *E. coli* population (Table 4).

In general, modeling errors in the two optimization approaches (GA and NSGA-II), based on MAPE and MAE in the train and validation datasets pinpointed that

NSGA-II-LSSVM to be the most accurate approach for the bacterial growth prediction in this study. The proximity between predicted and the observed values displayed the accuracy of the applied model (Fig. 3-5, Supplementary File 1 and Table 4). The errors of predictive models of the test dataset are also shown in test column of Table 4.

## 5. Discussion

In the present study, sigmoid functions and LSSVM based algorithms were employed to model bacterial growth curve. Unlike conventional LSSVM model, a MOO method (NSGA-II) was applied to calculate coefficients ('*c*' and '*σ*') for LSSVM training procedure. The accuracy of the proposed hybrid model was verified through several comparisons of the methods using MAE and MAPE.

Even though Logistics and Gompertz functions have widely been applied for modeling of the bacterial growth due to their simplicity (19, 37), they are less accurate comparing to the sophisticated approaches. Actually, nonlinear time series models, such as LSSVM, have been developed, because of limitations of the linear methods, to improve the forecasting performance (27). Consistent with our previous study (30), GA, due to its simplicity and accuracy, is a more appropriate parameter optimizer approach to calculate sigmoid function coefficients. However, to solve the optimization problems with two fitness functions, MOO approaches are superior (31). In this study, using NSGA-II for calculating the two coefficients: '*c*' and '*σ*', has resulted in a better optimization in addition to an increased accuracy of the bacterial growth modeling, accordingly. The hybrid models overcome the deficiencies of the individual models through merging different methods which result in the improvement of the prediction accuracy (2). The distinct differences between conventional LSSVM models and hybrid LSSVM model, aligned with various studies (1, 22, 26, 38), revealed the accuracy of hybrid models for the bacterial growth prediction in comparison to the single models. Although the developer of a prediction method should determine a proper model considering the characteristics of each method (2), finding the best scenario also depends on the nature of underlying data distribution, for instance, combination of feature selection (FS) and LSSVM has shown100% success rate in recognizing *E. coli* promoter gene sequence (22), or self-organizing map (SOM), and LSSVM has shown a promising alternative technique for river flow time series forecasting (1).

In conclusion, through exploring the capability and effectiveness of the idea of the hybrid modeling, we found NSGA-II-LSSVM outperforms other models, which in turn provides a promising alternative technique for prediction of the bacterial growth curve. In fact, the proposed model offers a better prediction due to its capability for fitting appropriate coefficients values. Prior studies, with different prediction approaches, have addressed other bacterial population datasets,

therefore, more efforts should be made using other strains, different datasets of experimental cases and/or under various environmental conditions to check the robustness of our proposed hybrid model. Additionally, finding a more optimal kernel function for a given learning task is still open to debate.

## Supplementary Files
This manuscript includes supplementary files accessible on http://www.ijbiotech.com.

## References
1. Ismail S, Shabri A, Samsudin R. A hybrid model of self organizing maps and least square support vector machine for river flow forecasting. *Hydrol Earth Syst Sci*. 2012;**16**(11):4417. doi: 10 10.5194/hess-16-4417-2012
2. Sallehuddin R, Shamsuddin SMH, Hashim SZM, Abraham A. Forecasting time series data using hybrid grey relational artificial neural network and auto regressive integrated moving average model. *Neural Network World*. 2007;**17**(6):573. doi:10.1080/08839510902879384
3. Chen Z, Li J, Wei L, Xu W, Shi Y. Multiple-kernel SVM based multiple-task oriented data mining system for gene expression data analysis. *Expert Syst Appl*. 2011;**38**(10):12151-12159. doi:10.1016/j.eswa.2011.03.025
4. Pang S, Ban T, Kadobayashi Y, Kasabov N. Personalized mode transductive spanning SVM classification tree. Inf Sci. 2011;**181**(11):2071-2085. doi.org/10.1016/j.ins.2011.01.008
5. Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*. 2003;**50**:159-175. doi:10.1016/S0925-2312(01)00702-0
6. Pouillot R, Lubran MB. Predictive microbiology models vs. modeling microbial growth within Listeria monocytogenes risk assessment: what parameters matter and why. *Food Microbiol*. 2011;**28**(4):720-726. doi:10.1016/j.fm.2010.06.002
7. Chatterjee T, Chatterjee BK, Majumdar D, Chakrabarti P. Antibacterial effect of silver nanoparticles and the modeling of bacterial growth kinetics using a modified Gompertz model. *Biochim Biophys Acta*. 2015;**1850**(2):299-306. doi.org/10.1016/j.bbagen.2014.10.022
8. Bernaerts K, Dens E, Vereecken K, Geeraerd AH, Standaert AR, Devlieghere F, et al. Concepts and tools for predictive modeling of microbial dynamics. *J Food Prot*. 2004;**67**(9):2041-2052. doi.org/10.4018/978-1-59904-945-8.ch044
9. Guillier L, Augustin J-C. Modelling the individual cell lag time distributions of Listeria monocytogenes as a function of the physiological state and the growth conditions. Int *J Food Microbiol*. 2006;**111**(3):241-251. doi:10.1016/j.ijfoodmicro.2006.05.011
10. Standaert AR, Francois K, Devlieghere F, Debevere J, Van Impe JF, Geeraerd AH. Modeling individual cell lag time distributions for Listeria monocytogenes. *Risk Anal*. 2007;**27**(1):241-254. doi.org/ 10.1111/j.1539-6924.2006.00873.x
11. Augustin JC, Zuliani V, Cornu M, Guillier L. Growth rate and

growth probability of Listeria monocytogenes in dairy, meat and seafood products in suboptimal conditions. *J Appl Microbiol.* 2005;**99**(5):1019-1042. doi:10.1111/j.1365-2672.2005.02710.x

12. Mejlholm O, Dalgaard P. Development and validation of an extensive growth and growth boundary model for Listeria monocytogenes in lightly preserved and ready-to-eat shrimp. *J Food Prot*. 2009;**72**(10):2132-2143. doi:10.4315/0362-028X-72.10.2132

13. Ross T, Rasmussen S, Fazil A, Paoli G, Sumner J. Quantitative risk assessment of Listeria monocytogenes in ready-to-eat meats in Australia. *Int J Food Microbiol.* 2009;**131**(2):128-137. doi:10.1016/j.ijfoodmicro.2009.02.007

14. Baty F, Delignette-Muller M-L. Estimating the bacterial lag time: which model, which precision? *Int J Food Microbiol*. 2004;**91**(3):261-277. doi:10.1016/j.ijfoodmicro.2003.07.002

15. 15. Dubey R, Maheshwari D. *Practical microbiology*: S. Chand Limited; 2002. 10.4172/2157-7471.1000385

16. Pommerville JC. *Alcamo's Fundamentals of Microbiology: Body Systems*: Jones & Bartlett Learning; 2009. doi: 10.1007/978-0-387-22458-9_3

17. Membré JM, Ross T, McMeekin T. Behaviour of Listeria monocytogenes under combined chilling processes. *Lett Appl Microbiol*. 1999;**28**(3):216-220. doi:10.1046/j.1365-2672.1999.00499.x

18. Augustin J-C, Brouillaud-Delattre A, Rosso L, Carlier V. Significance of Inoculum Size in the Lag Time of Listeria monocytogenes. *Appl Environ Microbiol*. 2000;**66**(4):1706-1710. doi.org/10.1128/AEM.66.4.1706-1710.2000

19. Perni S, Andrew PW, Shama G. Estimating the maximum growth rate from microbial growth curves: definition is everything. *Food Microbiol*. 2005;**22**(6):491-495. doi:10.1016/j.fm.2004.11.014

20. Suykens JA, Van Gestel T, De Moor B, Vandewalle J. *Least squares support vector machines:* World Scientific; 2002.

21. Polat K, Güneş S. Breast cancer diagnosis using least square support vector machine. *Digit Signal Process*. 2007;**17**(4):694-701. doi:10.1016/j.dsp.2006.10.008

22. Polat K, Güneş S, Arslan A. A cascade learning system for classification of diabetes disease: Generalized discriminant analysis and least square support vector machine. *Expert Syst Appl*. 2008;**34**(1):482-487. doi:10.1016/j.eswa.2006.09.012

23. Khemchandani R, Chandra S. Regularized least squares fuzzy support vector regression for financial time series forecasting. *Expert Syst Appl*. 2009;**36**(1):132-138. doi:10.1016/j.eswa.2007.09.035

24. Gencoglu MT, Uyar M. Prediction of flashover voltage of insulators using least squares support vector machines. *Expert Syst Appl*. 2009;**36**(7):10789-10798. doi:10.1016/j.eswa.2009.02.021

25. Esen H, Ozgen F, Esen M, Sengur A. Modelling of a new solar air heater through least-squares support vector machines. *Expert Syst Appl*. 2009;**36**(7):10673-10682. doi:10.1016/j.eswa.2009.02.045

26. Zhang K, Fei M, Li X, Zhou H. Adaptive bacteria colony picking in unstructured environments using intensity histogram and unascertained LS-SVM classifier. *Scientific World J.* 2014. doi:10.1155/2014/928395

27. Ma S, Lv M, Deng F, Zhang X, Zhai H, Lv W. Predicting the ecotoxicity of ionic liquids towards Vibrio fischeri using genetic function approximation and least squares support vector machine. *J Hazard Mater*. 2015;**283**:591-598. doi:10.1016/j.jhazmat.2014.10.011

28. García-Gimeno RMa, Hervás-Martínez C, de Silóniz MI. Improving artificial neural networks with a pruning methodology and genetic algorithms for their application in microbial growth prediction in food. *Int J Food Microbiol*. 2002;**72**(1):19-30. doi:10.1016/S0168-1605(01)00608-0

29. Kikuchi S, Tominaga D, Arita M, Takahashi K, Tomita M. Dynamic modeling of genetic networks using genetic algorithm and S-system. *Bioinformatics*. 2003;**19**(5):643-650. doi:10.1093/bioinformatics/btg027

30. Salehi Borojeni M, ShahMansourian A, Ghanegolmohammadi F, Nateghi A. Genetic algorithm is a more accurate approach to estimate maximum growth rate of bacteria growth pattern. 2nd *International Conferences on Mathematical Sciences (ICMS); Istanbul, Turkey* 2014.

31. Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput*. 2002;**6**(2):182-197. doi.org/10.1109/4235.996017

32. Moré JJ, Sorensen DC. Computing a trust region step. *SIAM J Sci Comput*. 1983;4(3):553-572. doi:10.1137/0904038

33. Zhang J, Tan Z, Yang S. Day-ahead electricity price forecasting by a new hybrid method. *Comput Ind Eng*. 2012;**63**(3):695-701. doi.org/10.1016/j.cie.2012.03.016

34. Hsu C-W, Chang C-C, Lin C-J. *A practical guide to support vector classification*. Last updated: 2016. Technical Report. University of National Taiwan, Taipei

35. Zhang W, Li C, Zhong B, editors. LSSVM parameters optimizing and non-linear system prediction based on cross validation. *Fifth International Conference on Natural Computation*; 2009: IEEE. doi:10.1109/ICNC.2009.26

36. Bigdeli N, Afshar K, Fotuhi-Firuzabad M. Bidding strategy in pay-as-bid markets based on supplier-market interaction analysis. *Energ Convers Manage*. 2010;**51**(12):2419-2430. doi:10.1016/j.enconman.2010.05.006

37. Zwietering M, Jongenburger I, Rombouts F, Van't Riet K. Modeling of the bacterial growth curve. *Appl Environ Microbiol*. 1990;**56**(6):1875-1881.doi:10.0099-2240/90/061875-07$02.00/0

38. Shi Z-b, Li Y, Yu T, editors. Short-term load forecasting based on LS-SVM optimized by bacterial colony chemotaxis algorithm. *International Conference on Information and Multimedia Technology*, 2009 ICIMT'09; 2009: IEEE. doi:10.1109/ICIMT.2009.57