Mühendislik Bilimleri ve Tasarım Dergisi
13(4), 1178-1787, 2025
e-ISSN: 1308-6693

Journal of Engineering Sciences and Design
DOI: 10.21923/jesd.1778201

Araştırma Makalesi
Research Article

PRECISION GROWTH MODELING OF ESCHERICHIA COLI IN SPARSE DATA SCENARIOS: A MULTI-STAGE LEARNING APPROACH

Hamit ARMAĞAN[1], Ulaş YAMANCI[2*],

[1], Department of Informatics, Suleyman Demirel University Isparta, Türkiye
[2] Faculty of Engineering and Natural Sciences Department of , Statistics, Suleyman Demirel University Isparta, Türkiye

Keywords
Microbial growth modeling,
Data augmentation,
Machine learning,
Regression model.

Abstract

This study investigates the prediction of *Escherichia coli* growth in environments where experimental data are limited, by integrating mathematical curve fitting with machine learning regression models. Two hybrid frameworks are developed: Fourier Series Curve Fitting combined with Gaussian Process Regression (FSCF-GPR), and Gaussian Curve Fitting integrated with Support Vector Machine Regression (GCF-SVMR). The raw dataset, initially composed of only 10 experimental measurements, was expanded to 114 data points through mathematical smoothing, providing a richer basis for model training. Model performance was assessed using Root Mean Square Error (RMSE), Mean Squared Error (MSE), Coefficient of Determination (R^2), and Mean Absolute Error (MAE). Results demonstrate that the FSCF-GPR framework achieved outstanding predictive accuracy with an R^2 of 0.9999, while GCF-SVMR also showed strong performance with an R^2 of 0.9934. These findings highlight that data augmentation via curve fitting can substantially enhance the accuracy and robustness of machine learning approaches in microbiological growth prediction under data-scarce conditions.

SEYREK VERİ SENARYOLARINDA ESCHERICHIA COLI'NİN HASSAS BÜYÜME MODELLEMESİ: ÇOK AŞAMALI BİR ÖĞRENME YAKLAŞIMI

Anahtar Kelimeler
Mikrobiyal büyüme modellemesi,
Veri artırma,
Makine öğrenmesi,
Regresyon model.

Öz

Bu çalışma, matematiksel eğri uydurma ile makine öğrenimi regresyon modellerini entegre ederek, deneysel verilerin sınırlı olduğu ortamlarda Escherichia coli büyümesinin tahminini araştırmaktadır. İki hibrit çerçeve geliştirilmiştir: Gauss Süreci Regresyonu ile birleştirilmiş Fourier Serisi Eğri Uydurma (FSCF-GPR) ve Destek Vektör Makinesi Regresyonu ile entegre edilmiş Gauss Eğri Uydurma (GCF-SVMR). Başlangıçta sadece 10 deneysel ölçümden oluşan ham veri kümesi, matematiksel düzeltme yoluyla 114 veri noktasına genişletilerek model eğitimi için daha zengin bir temel sağlanmıştır. Model performansı, Kök Ortalama Kare Hatası (RMSE), Ortalama Kare Hatası (MSE), Belirleme Katsayısı (R^2) ve Ortalama Mutlak Hata (MAE) kullanılarak değerlendirilmiştir. Sonuçlar, FSCF-GPR çerçevesinin 0,9999 R^2 ile olağanüstü bir tahmin doğruluğu elde ettiğini, GCF-SVMR'nin ise 0,9934 R^2 ile güçlü bir performans gösterdiğini ortaya koymaktadır. Bu bulgular, eğri uydurma yoluyla veri artırmanın, veri kıtlığı koşullarında mikrobiyolojik büyüme tahmininde makine öğrenimi yaklaşımlarının doğruluğunu ve sağlamlığını önemli ölçüde artırabileceğini vurgulamaktadır.

Alıntı / Cite
Armağan, H., Yamancı, U., (2025). Precision Growth Modeling of *Escherichia coli* in Sparse Data Scenarios: A Multi-Stage Learning Approach, Mühendislik Bilimleri ve Tasarım Dergisi, 13(4), 1178-1187.

Yazar Kimliği / Author ID (ORCID Number)	Makale Süreci / Article Process	
H. Armağan, 0000-0002-8948-1546	**Başvuru Tarihi / Submission Date**	02.10.2025
U. Yamancı, 0000-0002-4709-0993	**Revizyon Tarihi / Revision Date**	18.11.2025
	Kabul Tarihi / Accepted Date	20.11.2025
	Yayım Tarihi / Published Date	30.12.2025

---

[*] İlgili yazar / Corresponding author: ulasyamanci@sdu.edu.tr, +90-246-211-4352

PRECISION GROWTH MODELING OF ESCHERICHIA COLI IN SPARSE DATA SCENARIOS: A MULTI-STAGE LEARNING APPROACH

Hamit Armağan[1]†, Ulaş Yamancı,
[1], Department of Informatics, Suleyman Demirel University Isparta, Türkiye
[2] Faculty of Engineering and Natural Sciences Department of , Statistics, Suleyman Demirel University Isparta, Türkiye

Highlights

- Hybrid frameworks integrate curve fitting and machine learning
- FSCF-GPR framework achieved extremely high predictive accuracy
- Curve fitting-based data augmentation improves generalization

Purpose and Scope

This study aims to address the challenge of predicting microbial growth under sparse data conditions by integrating mathematical curve fitting with advanced machine learning methods.

Design/methodology/approach

Two hybrid frameworks are proposed. The first combines Fourier Series Curve Fitting with Gaussian Process Regression (FSCF-GPR), while the second integrates Gaussian Curve Fitting with Support Vector Machine Regression (GCF-SVMR). Data augmentation is performed through curve fitting, expanding the dataset from 10 to 114 points, followed by regression modeling.

Findings

The FSCF-GPR model achieved an R^2 value of 0.9999 with an RMSE of 0.00096 on the test set. The GCF-SVMR framework also yielded high accuracy ($R^2 = 0.9934$, RMSE = 0.0284). These results demonstrate that data augmentation through curve fitting significantly enhances the predictive power of machine learning models for microbial growth estimation

Research limitations/implications

The study is based on a limited dataset of 10 original measurements, augmented mathematically. Future studies could incorporate larger experimental datasets and evaluate different microbial species or environmental conditions.

Practical implications

The frameworks can be used by microbiologists and food safety experts to improve growth predictions without requiring extensive data collection, reducing experimental costs

Originality

This work introduces a novel combination of mathematical curve fitting and machine learning for microbial growth prediction in data-scarce environments. The proposed FSCF-GPR model outperforms many state-of-the-art approaches, including deep learning, under limited data conditions.

---

† İlgili yazar / Corresponding author: ulasyamanci@sdu.edu.tr, +90-246-211-4352

1. Introduction

Predicting microbial growth with accuracy is crucial for applications spanning food microbiology, industrial fermentation, and public health surveillance. *Escherichia coli* (E. coli), often employed as a model organism in microbiological studies, continues to serve as a reference point in both theoretical and applied research. Despite decades of work, forecasting its growth remains a nontrivial problem due to the complex dynamics and high sensitivity of biological systems to environmental factors (Baranyi and Roberts, 1994; Berwald vd., 2012; Cheroutre-Vialette and A. Lebert, 2002; Di Sciascio and Amicarelli, 2008).

Classical approaches, such as logistic, Gompertz, and Baranyi models, offer useful frameworks grounded in biological theory. However, these models often rely on stringent assumptions, and their accuracy declines when applied to heterogeneous or limited datasets. Their performance is particularly constrained by the quality and granularity of input data, and by their inability to fully capture irregularities in empirical growth curves (Fujikawa vd., 2004; Geeraerd vd., 2000).

- They require high-quality, evenly distributed data to ensure reliable parameter estimation;
- Their predictive capacity diminishes when applied to novel or fluctuating environmental conditions;
- Subtle nonlinearities and interactions in empirical datasets are often insufficiently captured (Jeyamkondan vd., 2001; Shamsudin vd., 2017).

Recent advances in machine learning (ML) have provided alternative pathways for modeling microbial behavior. Data-driven methods such as Support Vector Machines (SVM), Gaussian Process Regression (GPR), and Artificial Neural Networks (ANN) enable flexible mapping of complex relationships without the need for mechanistic assumptions (Serhat vd., 2024; Chitra vd., 2021). These models offer advantages including:

- The ability to learn directly from empirical observations;
- Enhanced adaptability across heterogeneous datasets;
- Tolerance for noise and variability inherent in microbiological data (Rizki vd., 2022; Juneja vd., 2019).

However, the use of ML in microbiology is not without caveats. Several practical limitations continue to affect implementation:

- Experimental datasets are often small, imbalanced, or sparse due to logistical and financial constraints (Choi vd., 2020);
- Overfitting becomes a risk when models are trained on limited or highly correlated data points;
- Interpretability of ML outputs may lag behind traditional parametric models, complicating biological validation (Jiang vd., 2022; Weller vd., 2021).

In response to these issues, this study proposes a multi-stage learning approach that seeks to combine thestructural strengths of mathematical modeling with the flexibility of machine learning. Specifically:

- The Fourier Series Curve Fitting with Gaussian Process Regression (FSCF-GPR) framework uses a truncated Fourier expansion to generate a high-resolution approximation of bacterial growth curves, which are subsequently used to train a GPR model;
- The Gaussian Curve Fitting with Support Vector Machine Regression (GCF-SVMR) framework employs Gaussian functions for data smoothing and augmentation prior to SVM-based regression modeling.

Both multi-stage frameworks aim to maximize the utility of limited datasets by first generating synthetic but biologically plausible growth patterns, followed by training of predictive ML models. This two-stage design improves generalization capacity while maintaining alignment with empirical observations.

The proposed methods are evaluated using standard performance metrics, including Root Mean Square Error (RMSE), Mean Squared Error (MSE), Coefficient of Determination (R^2), and Mean Absolute Error (MAE) (Ince vd., 2024; Kowalik vd., 2018). The goal is not only to assess predictive strength but also to offer a reproducible, resource-efficient methodology for microbial growth analysis under data-constrained conditions.

Overall, this work contributes to the growing body of research that integrates curve fitting, statistical learning, and biological modeling for enhanced interpretability and robustness in microbial growth forecasting.

2. Materials and Methods

2.1. Material and data collection

E. coli bacteria were grown in the laboratory at 37°C in a time-dependent manner in (Martinez-Rios vd., 2020) . This is the optimum growth temperature for E. coli. In such experiments, incubators are usually used and the growth of microbial organisms is measured and monitored at specific time intervals. In this way, the growth curve of the bacteria is constructed and the growth processes of E. coli are modeled using data from different time periods. During the experiment, the growth environment of the bacteria was controlled and provided with a nutrient-rich environment.

2.2. Fourier Series Curve Fitting and Gaussian Process Regression (FSCF-GPR) Framework

To address the challenges of limited experimental data in microbial growth modeling, we propose a hybrid methodology termed Fourier series curve fit-Gaussian Process Regression (FSCF-GPR). This method consists of two core stages: data augmentation using a truncated Fourier series, and predictive modeling using Gaussian Process Regression (GPR).

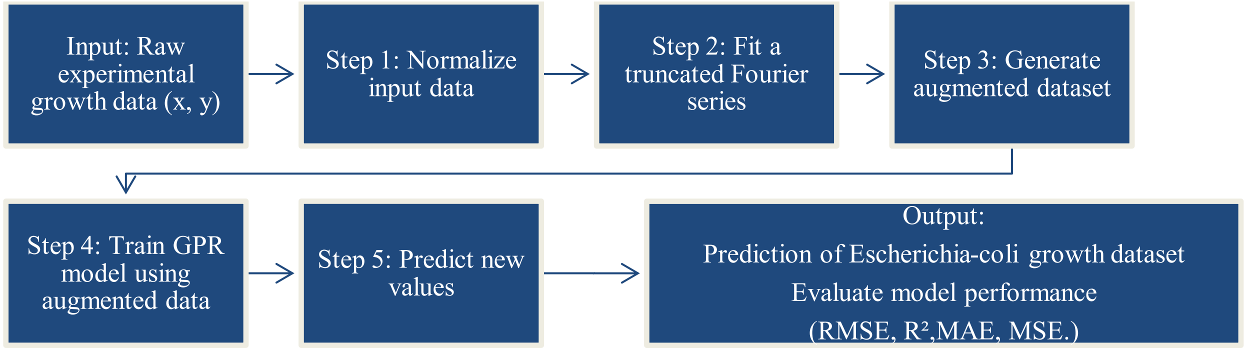

Figure 1. Workflow of the FSCF-GPR hybrid modeling framework.

The hybrid modeling process for predicting *Escherichia coli* growth using the FSCF-GPR framework consists of five systematic steps (figure 1).

- Step 1: Normalization: the input variable x is normalized using its mean and standard deviation to ensure numerical stability and scale invariance in the subsequent modeling process.
- Step 2: Fourier fitting: a third-order truncated Fourier series model is applied to the normalized data to capture both the general growth trend and potential oscillatory behavior of the microbial system.
- Step 3: Data Augmentation: based on the fitted Fourier curve, the original dataset, initially consisting of 10 observations, is expanded to 114 points to provide a denser and smoother representation of the growth pattern.
- Step 4: GPR Training: the augmented dataset is used to train a Gaussian Process Regression (GPR) model utilizing the Matérn 5/2 kernel, which offers flexibility in modeling moderately smooth nonlinear functions.
- Step 5: Prediction: the trained GPR model is then used to predict microbial growth on unseen test data, enabling high-precision forecasting even under limited original sample sizes.

In the first stage, an initial set of 10 experimental growth measurements of *E. coli* collected under controlled laboratory conditions (at 37°C, the organism's optimal growth temperature) is used to fit a truncated Fourier series (equation.1). This series is designed to capture both the overall growth trend and potential oscillatory behaviour in the microbial population. By applying the fitted Fourier model, the dataset is expanded to a total of 114 points, providing a smoother, high-resolution representation of bacterial growth over time.

$$\boldsymbol{f(x) = a_0 + a_1\cos(xw) + b_1\sin(xw) + a_2\cos(2xw) + b_2\sin(2xw) + a_3\cos(3xw) + b_3\sin(3xw)} \qquad (1)$$

In the second stage, the augmented dataset is used to train a GPR model employing the **Matérn 5/2 kernel** as the covariance function. This kernel is well-suited for modeling biological systems due to its flexibility and ability to handle non-linear, moderately smooth functions. The Matérn 5/2 kernel is showed as equation.2 (Palar vd., 2021).

$$k(x_i, x_j|\theta) = \sigma_f^2\left(1 + \frac{\sqrt{5}r}{\sigma_l} + \frac{5r^2}{3\sigma_l^2}\right) exp\left(-\frac{\sqrt{5}r}{\sigma_l}\right)$$

$$r = \sqrt{(x_i - x_j)^T(x_i - x_j)} \tag{2}$$

The Gaussian Process Regression (GPR) model employed in this study utilizes the Matérn 5/2 kernel function, characterized by hyperparameters sigmaL and sigmaF with estimated values of approximately 2.88 and 6.18, respectively. The model's regression coefficient (Beta) is -2.218, indicating the linear trend captured within the input space. The noise variance parameter (Sigma) is estimated at 0.00348, reflecting low observational noise in the data. Predictor normalization was performed with a location mean of 168.90 and a scale of 98.96 to ensure numerical stability. The active set size consists of 103 training observations, enabling effective learning and generalization.

2.3. Gaussian Curve Fitting and Support Vector Machine Regression (GCF-SVMR) Framework

As an alternative hybrid modeling strategy, we also develop the Gaussian Curve Fitting and Support Vector Machine Regression (GCF-SVMR) Framework. This method utilizes the same Fourier series-based data augmentation described in Section 2.2, followed by regression modeling via a Gaussian-kernel-based Support Vector Machine (SVM).

$$f(x) = a_1 e^{-\left(\frac{x-b_1}{c_1}\right)^2} + a_2 e^{-\left(\frac{x-b_2}{c_2}\right)^2} + a_3 e^{-\left(\frac{x-b_3}{c_3}\right)^2} \tag{3}$$

After the initial Gaussian Curve Fitting expands the dataset from 10 to 114 observations, the enhanced dataset is used to train an SVM regression model that employs the Radial Basis Function (RBF) kernel, also known as the Gaussian kernel. This kernel is defined as:

$$k(x_i, y_j) = exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \tag{4}$$

The parameter σ governs the kernel's flexibility: smaller values of σ allow the model to focus on local variations, while larger values generalize more broadly across the data. The RBF kernel enables the SVM to capture complex, non-linear relationships between the time variable and bacterial growth measurements with high precision.

By combining high-resolution data augmentation with kernel-based regression, the GCF-SVMR framework enhances predictive performance while maintaining robustness to noise and overfitting. This makes the model suitable for microbiological growth prediction where raw experimental data are scarce or unevenly distributed.

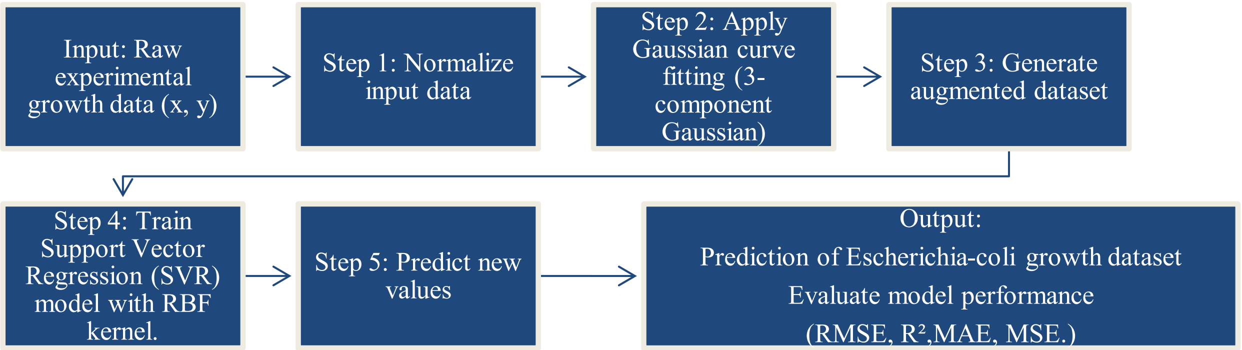

Figure 2. Workflow of the GCF-SVMR hybrid modeling framework.

To prepare the limited experimental data for predictive modeling, the raw growth measurements of E. coli—expressed as paired observations (x, y)—were first standardized. This normalization step, based on the sample mean (μ = 168.90) and standard deviation (σ = 98.96), was applied to reduce the influence of scale disparities and promote numerical stability throughout the training process.

After normalization, a three-component Gaussian fitting procedure was employed to approximate the underlying growth dynamics of the organism. This smoothed curve served not only to capture the essential structure of the biological trend but also to interpolate additional data points, expanding the original dataset from 10 to 114 observations. The augmented dataset thereby offered a denser and more continuous representation of the growth trajectory, enhancing the model's learning potential in a low-data regime.

Using this expanded dataset, a Support Vector Regression (SVR) model was trained, leveraging a Radial Basis Function (RBF) kernel. The choice of RBF was motivated by its capacity to represent complex, nonlinear relationships through implicit projection into a high-dimensional space. Model configuration involved careful tuning of hyperparameters to ensure accurate and generalizable predictions:

- The **ε-insensitive loss margin** was set to 0.0443, allowing small deviations from actual values without penalty, thereby improving tolerance to minor fluctuations in microbial growth.
- A **bias term** of approximately 2.0155 adjusted the model's intercept to align predictions with the central trend of the dataset.
- **Kernel scale** was fixed at 1, balancing the smoothness and sensitivity of the decision function to local data variations.

Following model training, predictions were generated and assessed using four statistical metrics widely used in regression analysis: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), and the Coefficient of Determination (R^2). These evaluations confirmed the SVR model's robustness and accuracy in capturing the nonlinear characteristics of *E. coli* growth dynamics, as visually summarized in Figure 2.

3. Experimental Results

3.1. FSCF-GPR Hybrid Framework

The growth data of *Escherichia coli* were modelled using a third-order Fourier series fit. The fitting results indicate good agreement between the model and observed data. Specifically, the Sum of Squares due to Error (SSE) was 0.0001, the coefficient of determination (R-square) was 0.9999, and the Adjusted R-square was 0.9997. The Root Mean Square Error (RMSE) was calculated as 0.0065, with a degree of freedom for error (DFE) of 2. These metrics collectively demonstrate that the Fourier model captures the underlying growth pattern of *E. coli* with high accuracy and minimal residual error.

Table.1. GPR regression model error and approximation values
(training data observations:103, predictors:1, validation: 5-fold cross-validation, test data observation:11)

Models	RMSE (Validation)	MSE (Validation)	RSquared (Validation)	MAE (Validation)
Matern 5/2 GPR	0.006070885	3.68556E-05	0.999695021	9.45288E-04
Models	RMSE (Test)	MSE (Test)	RSquared (Test)	MAE (Test)
Matern 5/2 GPR	0.000960846	9.23225E-07	0.999992412	3.03504E-04

Table 1 presents the error metrics and approximation values for the Gaussian Process Regression (GPR) model employing the Matern 5/2 kernel. The model was trained on 103 observations with a single predictor variable, validated using 5-fold cross-validation, and tested on an independent set of 11 observations. The validation results indicate a Root Mean Square Error (RMSE) of approximately 0.0061, a Mean Squared Error (MSE) of 3.69×10^{-5}, a high coefficient of determination (R^2) of 0.9997, and a Mean Absolute Error (MAE) below 0.001. On the test dataset, the model demonstrated even stronger performance, achieving an RMSE of 0.00096, an MSE of 9.23×10^{-7}, an R^2 value of 0.99999, and an MAE of 0.0003. These results collectively highlight the model's exceptional predictive accuracy and robustness in capturing the underlying microbial growth dynamics.

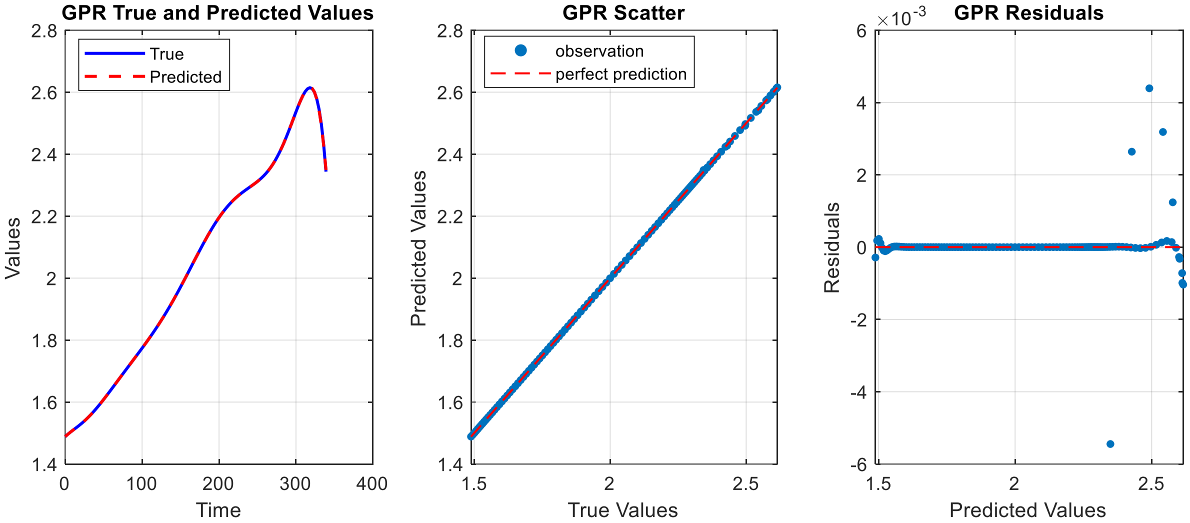

Figure.3. Predict GPR (Gaussian Process Regression) model analysis. (Left) Predict GPR model true and predicted values. (Center) GPR model scatter. (Right) GPR residuals.

Figure 3 illustrates the predictive performance of the Gaussian Process Regression (GPR) model using the Matern 5/2 kernel. The left panel displays the predicted values generated by the model, which closely align with the observed data, demonstrating the model's high approximation capability. The descriptive statistics of the target variable indicate a minimum value of 1.4886 and a maximum of 2.6156, with a mean of 2.0411 and a median of 2.0641, suggesting a near-symmetric distribution. The mode of 1.4886, along with a standard deviation of 0.3481 and a range of 1.1270, further underscores the moderate variability within the dataset. These results confirm that the GPR model not only captures the central tendency but also effectively models the spread and variation of microbial growth values under the given experimental conditions.

3.2. GCF-SVMR Hybrid Framework

The growth data of Escherichia coli were also modelled using a third-order Gaussian curve fitting function. Where the input variable x was normalized using a mean of 169 and a standard deviation of 100.9. The fitting performance demonstrates a highly accurate match between the model and the experimental data. The Sum of Squares due to Error (SSE) was 0.0001, the coefficient of determination (R-square) was 0.9999, and the Adjusted R-square was 0.9995. The Root Mean Square Error (RMSE) was calculated as 0.0080, with a degree of freedom for error (DFE) of 1. These results confirm that the Gaussian curve fitting model effectively captures the nonlinear behavior of bacterial growth with exceptional precision.

The performance of the Medium Gaussian SVM regression model was evaluated using both 5-fold cross-validation and a separate test dataset. During the validation phase, the model demonstrated strong predictive accuracy, with a Root Mean Squared Error (RMSE) of 0.0431, a Mean Squared Error (MSE) of 1.8582e-03, a high R-Squared value of 0.9846, and a Mean Absolute Error (MAE) of 0.0323. These metrics indicate that the model fits the training data well and is capable of making precise predictions. When applied to the test dataset, the Medium Gaussian SVM regression model continued to perform exceptionally, with improved metrics compared to the validation phase. The RMSE on the test data was 0.0284, the MSE was 8.0662e-04, the R-Squared value rose to 0.9934, and the MAE was 0.0236. These results confirm the model's robustness and its ability to generalize effectively to new, unseen data, show its potential for accurate and reliable predictions in practical applications (Table.2).

Table.2. SVM regression model error and approximation values
(training data observations:103, predictors:1, validation:5-fold cross-validation, test data observation:11)

Model	**RMSE**	**MSE**	**RSquared**	**MAE**
	(Validation)	**(Validation)**	**(Validation)**	**(Validation)**
Medium Gaussian SVM	0.043106647	1.858183E-03	0.984623618	0.032283849
Model	**RMSE**	**MSE**	**RSquared**	**MAE**
	(Test)	**(Test)**	**(Test)**	**(Test)**
Medium Gaussian SVM	0.028401129	8.066241E-04	0.993370784	0.02355325

The Support Vector Machine (SVM) regression model yielded highly accurate approximation results, reflected in its low prediction error metrics and strong generalization ability. The model effectively captured the underlying patterns in the data, with minimal deviation between predicted and actual values. Descriptive statistics of the target variable, which represents microbial growth, show that values ranged from 1.5392 to 2.5520, with a mean of 2.0502 and a median of 2.0705, indicating a near-symmetric distribution. The mode was observed at 1.5392, while the standard deviation was calculated as 0.3396, suggesting moderate variability in the dataset. The total range of 1.0128 highlights the scale of variation within the observed values. These results underscore the robustness of the SVM regression model in approximating microbial growth dynamics with high precision (figure.4).

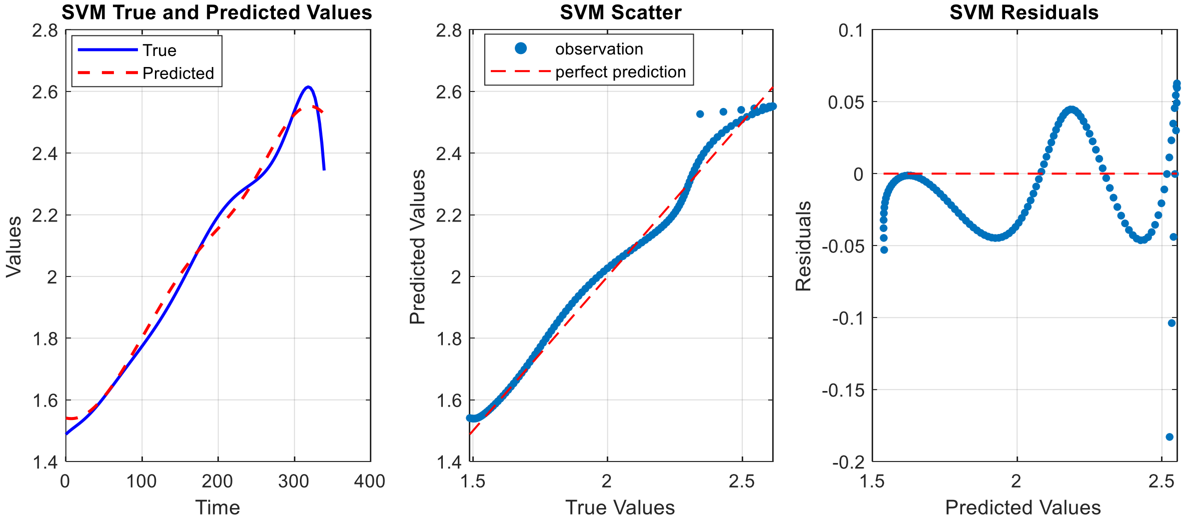

Figure.4. Predict SVM (support vector machine) model analysis. (Left) Predict SVM model true and predicted values. (Center) SVM model scatter. (Right) SVM residuals.

The predicted values generated by the Support Vector Machine (SVM) model exhibit a minimum of 1.5392 and a maximum of 2.5520, with a mean value of 2.0502. The median prediction closely aligns at 2.0705, while the mode corresponds to the minimum value of 1.5392, indicating a skew towards the lower bound in the predicted distribution. The standard deviation of 0.3396 reflects a moderate degree of variability within the predictions, and the range spans 1.0128, suggesting a reasonable spread in the model's output. These statistical descriptors collectively demonstrate the SVM model's capacity to capture the variability inherent in the data while maintaining consistency around the central tendency, as further illustrated in Figure 4, which compares the true and predicted values.

4. Discussion and Conclusion

Modeling *E. coli* growth with high fidelity remains difficult, especially under sparse data conditions. While various machine learning techniques have been adopted to address this challenge, their predictive success is often contingent on dataset characteristics, model tuning, and preprocessing strategies. Notably, deep learning methods such as NARX and LSTM have demonstrated excellent results when sufficient temporal resolution is available.

In contrast, our study focuses on a hybrid strategy that begins with data augmentation through curve fitting before engaging regression modeling. The proposed FSCF-GPR and GCF-SVMR frameworks show strong predictive performance, even with minimal original data. In particular, the FSCF-GPR model exhibits extremely low error margins, outperforming many established approaches in the literature.

In the literature, Shamsudin et al. (2017) reported highly accurate predictions of E. coli growth using a NARX neural network, achieving a regression value of 0.99956 with a relatively simple network structure consisting of 2 hidden neurons and 7 delays. Similarly, Chitra et al. (2021) compared deep learning architectures and showed that LSTM models outperformed others, yielding an R^2 value of 0.9994 and RMSE of 0.011. Their study confirmed that advanced temporal models such as NARX and LSTM can model microbial time-series data effectively when sufficient temporal resolution is available. In a separate study, Al et al. (2024) evaluated multiple machine learning algorithms for E. coli O157:H7 growth prediction in raw ground beef. Random Forest (RF) and Artificial Neural Networks (ANN) achieved high R^2 values of 0.98 and 0.96, respectively, while simpler models such as Multiple Linear Regression (MLR) performed significantly worse ($R^2 = 0.66$).

Compared to these existing approaches, the hybrid frameworks proposed in this study—Fourier Series Curve Fitting with Gaussian Process Regression (FSCF-GPR) and Gaussian Curve Fitting with Support Vector Machine Regression (GCF-SVMR)—demonstrate superior or comparable performance, particularly under sparse data conditions. The FSCF-GPR model achieves a test R^2 value of 0.9999 with a remarkably low RMSE of 0.00096, outperforming most previously published results including deep neural models. The GCF-SVMR model also yields strong predictive accuracy with a test R^2 of 0.9934 and RMSE of 0.0284, making it competitive with high-performance ANN and RF models from the literatüre (table.3).

Table 3. Comparative summary of recent *E. coli* growth modeling studies based on different machine learning and hybrid approaches.

Year	Author(s)	Model Type	R^2
2017	Shamsudin et al.	NARX ANN	$R^2 = 0.99956$
2021	Chitra et al.	LSTM	$R^2 = 0.9994$
2024	Al et al.	RF, ANN, MLR	RF: $R^2 = 0.98$ ANN: $R^2 = 0.96$ MLR: $R^2 = 0.66$
2025	Armağan & Yamancı	FSCF-GPR (Fourier + Gaussian Process)	$R^2 = 0.9999$
2025	Armağan & Yamancı	GCF-SVMR (Gaussian Fit + SVM Regression)	$R^2 = 0.9934$

The key advantage of the proposed hybrid approaches lies in their integration of mathematical curve fitting for data augmentation and smoothing, followed by training of advanced machine learning models on the augmented datasets. This preprocessing step not only improves model generalization but also reduces the need for large-scale data collection, which is particularly beneficial in microbiological studies where experiments are often resource-intensive.

In conclusion, this study introduces two effective hybrid methodologies—FSCF-GPR and GCF-SVMR—for modeling microbial growth under data-limited conditions. By combining curve fitting with machine learning, the proposed methods achieve high prediction accuracy, outperforming many state-of-the-art models reported in the literature. These frameworks can serve as robust tools for microbiologists, biotechnologists, and food safety analysts seeking reliable predictive models in environments where data acquisition is constrained. Future research may explore the extension of these models to other microbial species, different environmental variables, or integration with mechanistic biological models for improved interpretability.

Acknowledgement

The authors would like to express their sincere gratitude to Selmihan Şahin ABDULMAJEED (Süleyman Demirel University, Department of Chemistry) for their valuable contributions and insightful support throughout the study. The authors also would like to sincerely thank the anonymous referees for some useful comments and suggestions.

Conflict of Interest

No conflict of interest was declared by the authors.

References

Al, S., Uysal Ciloglu, F., Akcay, A., Koluman, A., 2024. Machine learning models for prediction of Escherichia coli O157: H7 growth in raw ground beef at different storage temperatures, Meat Science, 210, 1-7.

Baranyi, J., Buss da Silva, N., and Ellouze, M., 2017. Rethinking tertiary models: relationships between growth parameters of Bacillus cereus strains. Front. Food Microbiol., 8, 1-8.

Baranyi, J., Roberts, T. A., 1994. A dynamic approach to predicting bacterial growth in food. International Journal of Food Microbiology, 23, 277--294.

Berwald, J., Gedeon, T., Sheppard, J., 2012Using machine learning to predict catastrophes in dynamical systems, J. Comput. Appl. Math. 236(9), 2235-2245.

Cheroutre-Vialette, M. , Lebert, A., 2002. Application of recurrent neural network to predict bacterial growth in dynamic conditions, International Journal of Food Microbiology, 73, 107-118.

Chitra, M., Sutha, S., Pappa, N., 2021. Application of deep neural techniques in predictive modelling for the estimation of Escherichia coli growth rate, Journal of Applied Microbiology, 130(5), 1645-1655.

Choi, M.S., Kim, J.Y., Jeon, E.B., Park, S.Y., 2020. Predictive growth models of Bacillus cereus on dried laver Pyropia pseudolinearis as a function of storage temperature. Korean Journal of Fisheries and Aquatic Sciences, 53(5), 699-706.

Di Sciascio, F., Amicarelli, A.N., 2008. Biomass estimation in batch biotechnological processes by Bayesian Gaussian process regression, Computers & Chemical Engineering, 32, 3264--327.

Fujikawa, H., Kai, A., Morozumi, S., 2004. A new logistic model for Escherichia coli growth at constant and dynamic temperatures, Food Microbiology, 21, 501-509.

Geeraerd, A.H., Herremans, C.H., Van Impe, J.F., 2000. Structural model requirements to describe microbial inactivation during a mild heat treatment, International Journal of Food Microbiology, 59 185-209.

Ince, V., Bader-El-Den, M., Arabikhan, F., Sari, O. F., 2024. Machine Learning for Bacterial Growth Prediction and Examination of Bacterial Impact on Growth, In 2024 IEEE 12th International Conference on Intelligent Systems (IS), 1-6.

Jeyamkondan, S., Jayas, D.S., Holley, R.A., 2001.Microbial growth modelling with artificial neural networks. International Journal of Food Microbiology, 64 343-354.

Jiang, Y.N., Luo, J., Huang, D., Liu, Y., Li, D., 2022. Machine learning advances in microbiology: A review of methods and applications. Frontiers in Microbiology, 13, Article 925454.

Juneja, V.K., Golden, C. E., Mishra, A., Harrison, M., Mohr T.B., 2019. Predictive model for growth of Bacillus cereus at temperatures applicable to cooling of cooked pasta, Journal of Food Science , 84(3), 590-598.

Kowalik, J., Lobacz, A., Zulewska, J., Dec, B., 2018. Analysis and mathematical modelling of the behaviour of Escherichia coli in the mascarpone cheese during cold storage. International Journal of Food Science and Technology, 53(6), 1541–1548.

Martinez-Rios, V., Gkogka, E., Dalgaard, P., 2020. Predicting growth of Listeria monocytogenes at dynamic conditions during manufacturing, ripening and storage of cheese - Evaluation and application of models. Food Microbiology, 92, 1–12.

Palar, P.S., Zakaria, K, Zuhal, L.R., Shimoyama, K., Liem, R.P., 2021. Gaussian processes and support vector regression for uncertainty quantification in aerodynamics. AIAA Scitech 2021 Forum (AIAA, 2021), p. 0181.

Rizki, W. M., Ratih D. H, Harsi D.K, 2022. Comparison of Predictive Growth Models for Bacillus Cereus in Cooked and Fried Rice During Storage, The Annals of the University Dunarea De Jos of Galati. Fascicle VI - Food Technology, 46 (2), 89-103.

Shamsudin, S.N., Rahiman, M.H.F., Taib, M.N, Ahmad, A.H, Razak, R.W.A., 2017 Escherichia coli growth modeling using neural network. J. Fundam. Appl. Sci., 9, 759-771.

Stavropoulou, E. and Bezirtzoglou, E., 2019. Predictive modeling of microbial behavior in food. Foods, 8(12) 1-16.

Weller, D., Love, T., Wiedmann M., (2021). Interpretability versus accuracy: A comparison of machine learning models built using different algorithms, performance measures, and features to predict E. coli levels in agricultural water, Frontiers in Artificial Intelligence, 4, Article 628441.